



Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ingeniería de Sistemas e Informática
Unidad de Posgrado

**Gestión de datos de investigación en universidades en
base al ciclo de vida de los datos. Caso de estudio: Área
de Ciencias de la Salud**

TESIS

Para optar el Grado Académico de Magíster en Gestión de la
Información y del Conocimiento

AUTOR

María del Pilar Liliana ELÉSPURU BRICEÑO

ASESOR

Nora Bertha LA SERNA PALOMINO

Lima, Perú

2017



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Eléspuru, M. (2017). *Gestión de datos de investigación en universidades en base al ciclo de vida de los datos. Caso de estudio: Área de Ciencias de la Salud*. [Tesis de maestría, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática, Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
 Universidad del Perú, DECANA DE AMÉRICA
Facultad de Ingeniería de Sistemas e Informática
UNIDAD DE POSGRADO

**SUSTENTACIÓN DE TESIS PARA OPTAR EL GRADO ACADÉMICO DE MAGÍSTER
 EN GESTIÓN DE LA INFORMACIÓN Y DEL CONOCIMIENTO**

En la Ciudad Universitaria, a los dos (02) días del mes de junio del 2017, siendo las 18:30 horas, se reunieron en el Aula Magna de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional Mayor de San Marcos, el Jurado Examinador de tesis conformado por los siguientes profesores:

Mg. Zoraida Emperatriz Mamani Rodríguez (Presidente)
 Dra. Nora Bertha La Serna Palomino (Miembro Asesor)
 Mg. Alfredo Celso Alva Bravo (Miembro)
 Mg. Sergio Valcarcel Ascencios (Miembro)
 Mg. Víctor Darío Miranda Vargas (Miembro)

Se inició la Sustentación de la tesis invitando a la graduando **María Del Pilar Liliana Eléspuru Briceño**, para que realizara la exposición oral y pública de la tesis para optar el Grado Académico de Magister en Gestión de la Información y del Conocimiento, siendo la Tesis intitulada:

**“Gestión de Datos de Investigación en Universidades en Base al Ciclo de Vida de los Datos.
 Caso de Estudio: Área de Ciencias de la Salud”**

Concluida la exposición, los miembros del Jurado Examinador procedieron a formular sus preguntas que fueron absueltas por el graduando; acto seguido se procedió a la evaluación correspondiente, habiendo obtenido la siguiente calificación:

..... DIECISIETE (17) MUY BUENO

Por tanto el Presidente del Jurado, de acuerdo al Reglamento de Grados y Títulos, le otorga a la bachiller **María Del Pilar Liliana Eléspuru Briceño** el Grado Académico de Magister en Gestión de la Información y del Conocimiento, cuyo expediente debe ser remitido al Consejo de Facultad para su aprobación.

Siendo las 19:30 horas, el Presidente del Jurado Examinador da por concluido el acto académico de Sustentación de Tesis.

MG. ZORAIDA EMPERATRIZ MAMANI RODRIGUEZ
 Presidente

DRA. NORA BERTHA LA SERNA PALOMINO
 Miembro Asesor

MG. ALFREDO CELSO ALVA BRAVO
 Miembro

MG. SERGIO VALCARCEL ASCENCIOS
 Miembro

MG. VÍCTOR DARIO MIRANDA VARGAS
 Miembro

DEDICATORIA

A mi padre, quien aunque no esté presente, siempre está en mi mente y sé que hubiera estado muy feliz al verme graduada de Magíster, debido a que él siempre, con sus sabios consejos me animaba continuamente a superar los retos para lograr mis metas.

AGRADECIMIENTO

Deseo agradecer a mi familia quienes siempre me han apoyado a lo largo de esta investigación.

A Ernesto, mi esposo, quien ha tenido la paciencia de acompañarme los fines de semana. A mis hijos Ernesto y Giuliana, quienes siempre me han animado a seguir adelante y no rendirme.

A mis amigos, especialmente a Miguel Estrada, por haberme brindado y apoyado con su tiempo para resolver mis dudas y llevar adelante esta investigación.

Al equipo de Apoyo a la Producción Intelectual de la Dirección de Gestión del Conocimiento de la UPC, por escucharme y apoyado con mucho cariño para seguir adelante.

A mi asesora, la Dra. Nora La Serna Palomino, por su guía y por haberme brindado su tiempo para que esta investigación cumpla con los objetivos trazados.

A mi alma máter, la Universidad Nacional Mayor de San Marcos, por haberme brindado la oportunidad de cursar la Maestría en Gestión de Información y del Conocimiento en la Escuela de Postgrado de la Facultad de Ingeniería de Sistemas e Informática. De igual manera a la plana docente de postgrado.

RESUMEN

En este estudio encontramos, que los investigadores del Área de Ciencias de la Salud de la Universidad Nacional Mayor de San Marcos (UNMSM) y la Facultad de Ciencias de la Salud de la Universidad Peruana de Ciencias Aplicadas (UPC), no comparten o no dan acceso a sus datos de investigación a otros investigadores, salvo que se los soliciten de manera expresa primero porque tienen un gran sentido de propiedad y protección sobre los mismos y segundo porque desconocen de los beneficios de compartir y dar acceso a sus datos de investigación. Como explica Rogova, “al acceder a los datos hay que tener en cuenta que estos pueden ser imprecisos y que esta imprecisión se puede por que los datos pueden ser incompletos y ser pocos fiables o ambiguos debido a los errores de medición”. También tenemos que de los 25 investigadores encuestados (de la UNMSM y de la UPC), el 72% (18 investigadores) no reutilizan los datos de investigación de otros investigadores. Por otro lado, los investigadores de ambas universidades no cuentan con un estándar de registro y descripción de datos de investigación y que, por el contrario, utilizan medios y formatos muy diversos. Solo el 6% (4 investigadores), utilizan un repositorio de datos. En el contexto investigado, se hace evidente que los investigadores al no compartir sus datos, dificultan una gestión de datos orientada a mejorar la calidad y el número de investigaciones. En este sentido, lo que motivó la investigación fue desarrollar un modelo que satisfaga las expectativas de los investigadores en cuanto a la factibilidad, disponibilidad, uso de los datos de investigación, organización y almacenamiento, poniendo énfasis en la difusión sobre los beneficios de compartir y reutilizar los datos de investigación. A partir de este hallazgo, se plantea como principal agente de cambio, la gestión del ciclo de vida de los datos de investigación de manera sistemática, a través de un modelo que se ajuste a la realidad de la universidad peruana.

PALABRAS CLAVES

Gestión de datos de investigación; Acceso a los datos de investigación; Repositorios de datos; Facultades de Ciencias de la Salud.

ABSTRACT

In this study we find that, despite the willingness of researchers from the Health Sciences Area of the Universidad Nacional Mayor de San Marcos (UNMSM) and the School of Health Sciences at Universidad Peruana de Ciencias Aplicadas (UPC), do not share or do not give access to their research data to other researchers unless expressly requested. This happens because they have a great sense of ownership and protection over their work and secondly because they are unaware of the benefits of sharing and giving access to their research data. As Rogova (2010) explains, "when accessing the data we must take into account that these may be imprecise and that this vagueness can be because the data may be incomplete and be few reliable or ambiguous due to measurement errors ". So, from the 25 researchers surveyed (from UNMSM and UPC), 72% (18 researchers) do not re-use research data from other researchers. On the other hand, it is found that the researchers of both universities do not have a standard of registration and description of research data, on the contrary, they use very diverse resources and formats. Only 6% (4 researchers) use a data repository. In the context of the research, it is evident that researchers do not share their data, which makes it difficult to manage data aimed at improving the quality and number of investigations. In this sense, what motivated the research is to develop a model that satisfies the expectations of the researchers regarding the feasibility, availability, use of research data, organization and storage, with emphasis on the dissemination of the benefits of sharing and reuse research data. Based on this finding, the main agent of change is the management of the life cycle of research data in a systematic way, through a model that fits the reality of the Peruvian university.

KEYWORDS: Research data management; Research data access; Open Data; Data repositories; Health Sciences Areas.

INDICE GENERAL

INTRODUCCIÓN	1
1.1. Situación problemática.....	4
1.2. Formulación del problema principal.....	6
1.2.1. Problemas específicos.....	7
1.3. Justificación teórica	7
1.4. Justificación práctica	10
1.5. Objetivos de la investigación	11
1.5.1. Objetivo general.....	11
1.5.2. Objetivos específicos	11
1.6. Variables de estudio	12
1.6.1. Variable independiente	12
1.6.2. Variables dependientes	12
1.7. Operacionalización de variables	12
1.7.1. Variable dependiente 1	12
1.7.2. Variable dependiente 2	13
1.7.3. Variable dependiente 3	13
CAPÍTULO II: MARCO TEÓRICO.....	14
2.1. Introducción	14
2.2. Antecedentes de la investigación.....	14
2.3. Revisión Bibliográfica.....	20
2.3.1. Gestión de datos de investigación	20
2.3.2. Preservación de datos de investigación	34
2.3.3. Políticas y lineamientos para la gestión de los datos de investigación.....	36
2.3.3.1. Políticas y lineamientos para la gestión de datos de investigación a nivel internacional.....	36
2.3.3.2. Políticas y lineamientos para la gestión de datos de investigación en América Latina.....	42
2.3.3.3. Políticas y lineamientos para la gestión de los datos de investigación en el Perú.....	45
2.3.4. Estándares y lineamientos sobre el uso de softwares y metadatos a nivel internacional y nacional	54
2.3.4.1. Estándares de metadatos a nivel internacional y nacional	54
2.3.4.2. Estándares de softwares de datos de investigación a nivel internacional y nacional.....	57
2.3.4.2.1. Software CKAN.....	60
2.3.4.2.2. Software DKAN.....	60
2.3.5. Investigación en universidades: Facultades de Ciencias de la Salud	62
2.3.5.1. Investigación en las universidades en Perú.....	62
2.3.5.2. Facultades de Ciencias de la Salud	63
2.3.5.2.1. Ciencias de la Salud en la Universidad Nacional Mayor de San Marcos (UNMSM).	64
2.3.5.2.2. Facultad de Ciencias de la Salud en la Universidad Peruana de Ciencias Aplicadas UPC.....	66
2.4. Cuadro comparativo de Modelos del ciclo de vida de los datos y el modelo propuesto.....	71

CAPÍTULO III: MODELO PROPUESTO PARA LA GESTIÓN DE DATOS DE INVESTIGACIÓN EN BASE AL CICLO DE VIDA DE LOS DATOS	74
3.1. Mapa conceptual de la gestión de datos de investigación en base al ciclo de vida de los datos.....	74
3.2. Modelo Propuesto	77
3.3. Etapas del ciclo de vida de los datos del modelo propuesto (ver Tabla 11).	79
3.3.1. Creación de los datos de investigación	79
3.3.2. Evaluación, análisis y selección de los datos.	79
3.3.3. Preservación de los datos de investigación.....	80
3.3.4. Almacenamiento de los datos	80
3.3.5. Acceso y uso de los datos	81
3.3.6. Reutilización	81
3.3.7. Transformación de los datos	84
3.3.8. Difusión de los datos de investigación	84
CAPÍTULO IV: METODOLOGÍA	86
4.1. Tipo y Diseño de investigación	86
4.2. Matriz de operacionalización de variables	86
4.3. Unidad de análisis.....	88
4.4. Población	89
4.5. Muestra	89
4.6. Técnica de recolección de datos	89
4.7. Elaboración del cuestionario.....	90
CAPÍTULO V: ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS	97
5.1. Análisis e interpretación de los resultados UNMSM.....	97
5.2. Análisis e Interpretación de los resultados UPC	119
5.3. Análisis Comparativo e Interpretación de UNMSM – UPC	143
5.4. Conclusiones de la encuesta realizada en la UNMSM y la UPC.....	165
5.5. Beneficios que aporta la propuesta	168
CAPÍTULO VI: CONCLUSIONES GENERALES	169
6.1. Conclusiones respecto al Objetivo General	169
6.2. Conclusiones respecto a los Objetivos Específicos.....	170
REFERENCIAS BIBLIOGRÁFICAS	173

ÍNDICE DE GRÁFICOS

Gráfico 1 Agentes involucrados en la gestión de los datos de investigación	21
Gráfico 2 Ciclo de vida de los datos de investigación	23
Gráfico 3 Transformación de los datos de investigación	26
Gráfico 4 Plan de Gestión de datos de investigación.....	29
Gráfico 5 Etapas para el ciclo de vida de los datos según UD Data Archive	31
Gráfico 6: Portal de Datos Abiertos de la Unión Europea	40
Gráfico 7 Significado de Datos Abiertos	46
Gráfico 8 Portal Nacional de Datos Abiertos.....	48
Gráfico 9 Portal del Repositorio Digital ALICIA.....	49
Gráfico 10 Proporción de repositorios de datos por países según Open Doar.....	51
Gráfico 11 Tipos de contenidos de datos en repositorios según Open Doar	51
Gráfico 12 Tipos de softwares utilizados en repositorio de datos según Open Doar	52
Gráfico 13 Concytec. Portal de datos abiertos	53
Gráfico 14 Ministerio de Educación. Portal de Datos Abiertos.....	61
Gráfico 15 Ministerio de Salud. Portal de Datos Abiertos.....	61
Gráfico 16 Mapa conceptual de la gestión de datos de investigación en base al ciclo de vida de los datos	76
Gráfico 17 Modelo Propuesto para la Gestión de Datos de Investigación basado en el ciclo de vida de los datos.....	77
Gráfico 18 Frecuencia relativa: instituciones donde los investigadores de CS de la UNMSM realizan sus trabajos de investigación	98
Gráfico 19 Frecuencia relativa: líneas de investigación a las que se dedican los investigadores de CS de la UNMSM	99
Gráfico 20 Frecuencia relativa: las agencias que contribuyen con fondos a la investigación en UNMSM	99
Gráfico 21 Frecuencia relativa: uso que le da el investigador de CS de la UNMSM a sus de datos de investigación	100
Gráfico 22 Frecuencia relativa: tiempo que conserva el investigador de CS de la UNMSM sus datos de investigación	101
Gráfico 23 Frecuencia relativa: cantidad de datos de investigación que el investigador de la UNMSM pone a disponibilidad de los demás	102
Gráfico 24 Frecuencia relativa: usuarios que tiene acceso a los datos de investigación de los investigadores de CS de la UNMSM	103
Gráfico 25 Frecuencia relativa: medios a través de los cuales se accede a los datos de investigación de los investigadores de CS de la UNMSM.....	104
Gráfico 26 Frecuencia relativa: utilización de datos generados por otros investigadores o instituciones, por los investigadores de CS de la UNMSM.....	105
Gráfico 27 Frecuencia relativa: reutilización de datos generados por otros investigadores o instituciones, realizados por los investigadores de CS de la UNMSM.....	106

Gráfico 28 Participación del tipo de licencia que aplican los investigadores de CS de la UNMSM.....	106
Gráfico 29 Frecuencia relativa: sobre como registran sus datos de investigación los investigadores de CS de la UNMSM	107
Gráfico 30 Frecuencia relativa: tipo de formato que usa el investigador de CS de la UNMSM para almacenar sus datos de investigación.....	108
Gráfico 31 Frecuencia relativa: técnicas de recolección de datos que utiliza el investigador de CS de la UNMSM	109
Gráfico 32 Frecuencia relativa: tipo de datos de investigación que utilizan los investigadores de CS de la UNMSM	110
Gráfico 33 Frecuencia relativa: condiciones requeridas por los investigadores de CS de la UNMSM para enviar sus datos de investigación a un repositorio	111
Gráfico 34 Frecuencia relativa: espacio de almacenamiento que necesita los investigadores de CS de la UNMSM	113
Gráfico 35 Frecuencia relativa: tipo de almacenamiento que actualmente utiliza el investigador de CS de la UNMSM	114
Gráfico 36 Frecuencia relativa: de las razones más importantes para los investigadores de CS de la UNMSM para que sus datos de investigación estén disponibles y preservados	115
Gráfico 37 Frecuencia relativa: opciones de apoyo que el investigador de CS de la UNMSM esperaría de su institución si implementará un repositorio	117
Gráfico 38 Frecuencia relativa: preferencia del responsable para hacerse cargo del archivo de datos de investigación de los investigadores de CS de la UNMSM	118
Gráfico 39 Frecuencia relativa: instituciones donde los investigadores de la UPC realizan sus trabajos de investigación	119
Gráfico 40 Frecuencia relativa: líneas de investigación a las que se dedican los investigadores de la UPC	121
Gráfico 41 Frecuencia relativa: agencias que contribuyen con fondos a la investigación en UNMSM	122
Gráfico 42 Frecuencia relativa: uso que le da el investigador de la UPC a sus de datos de investigación	123
Gráfico 43 Frecuencia relativa: tiempo que conserva el investigador de la UPC sus datos de investigación	124
Gráfico 44 Frecuencia relativa: cantidad de datos de investigación que el investigador de la UPC pone a disponibilidad de los demás	125
Gráfico 45 Frecuencia relativa: usuarios que tiene acceso a los datos de investigación de los investigadores de CS de la UPC.....	126
Gráfico 46 Frecuencia relativa: medios a través de los cuales se accede a los datos de investigación de los investigadores de CS de la UPC	127
Gráfico 47 Frecuencia relativa: utilización de datos generados por otros investigadores o instituciones, por los investigadores de CS de la UPC	128
Gráfico 48 Frecuencia relativa: reutilización de datos generados por otros investigadores o instituciones, por los investigadores de CS de la UPC	129

Gráfico 49 Frecuencia relativa: tipo de licencia que aplican los investigadores de CS de la UPC	130
Gráfico 50 Frecuencia relativa: sobre como registra sus datos de investigación los investigadores de CS de la UPC	131
Gráfico 51 Frecuencia relativa: tipo de formato que usa el investigador de CS de la UPC para almacenar sus datos de investigación	132
Gráfico 52 Frecuencia relativa: técnicas de recolección de datos que utiliza el investigador de CS de la UPC.....	133
Gráfico 53 Frecuencia relativa: tipo de datos de investigación que utilizan los investigadores de CS de la UPC	134
Gráfico 54 Frecuencia relativa: condiciones requeridas por los investigadores de CS de la UNMSM para enviar sus datos de investigación a un repositorio	136
Gráfico 55 Frecuencia relativa: espacio de almacenamiento que necesitan los investigadores de CS de la UPC	137
Gráfico 56 Frecuencia relativa: tipo de almacenamiento que actualmente utiliza el investigador de CS de la UPC.....	138
Gráfico 57 Frecuencia relativa: razones más importantes para los investigadores de CS de la UPC para que sus datos de investigación estén disponibles y preservados.....	139
Gráfico 58 Frecuencia relativa: opciones de apoyo que el investigador de CS de la UPC esperaría de su institución si implementará un repositorio	141
Gráfico 59 Frecuencia relativa de la preferencia del responsable para hacerse cargo del archivo de datos de investigación de los investigadores de CS de la UPC.....	142

ÍNDICE DE TABLAS

Tabla 1: Acciones aplicables a todo el ciclo de vida de los datos.....	24
Tabla 2: Acciones secuenciales.....	24
Tabla 3: Acciones ocasionales	25
Tabla 4: Cuadro comparativo del ciclo de vida de los datos: DDC y UD Data Archive.....	32
Tabla 5 Clasificación de datos de investigación según la Research Information Network (RIN)	34
Tabla 6: Repositorios específicos para publicaciones.....	38
Tabla 7 Licencias compatibles con los principios de datos abiertos.....	59
Tabla 8 Formatos frecuentemente utilizados para la apertura de la información	59
Tabla 9 Características de la educación Médica en el Perú para el 2005	68
Tabla 10 Análisis Comparativo de Modelos del Ciclo de vida de los datos implementados con modelo propuesto.....	73
Tabla 11 Etapas del modelo propuesto basado en el ciclo de vida de los datos de investigación	78
Tabla 12 Comparación de esquemas de metadatos más utilizados para la gestión de datos de investigación	82
Tabla 13 Formatos recomendados para la reutilización de datos de investigación ..	83
Tabla 14 Licencias compatibles con los principios de datos abiertos.....	83
Tabla 15 Comparación de softwares de código abierto para la publicación de datos abiertos	84
Tabla 16 Matriz de operacionalización de variables.....	87
Tabla 17 Distribución de frecuencias: instituciones donde los investigadores de CS de la UNMSM realizan sus trabajos de investigación	97
Tabla 18 Distribución de frecuencias: líneas de investigación a las que se dedican los investigadores de CS de la UNMSM.....	98
Tabla 19 Distribución de frecuencias: agencias que contribuyen con fondos a la investigación en UNMSM	99
Tabla 20 Distribución de frecuencias: uso que le da el investigador de CS de la UNMSM a sus datos investigación	100
Tabla 21 Distribución de frecuencias: tiempo que conserva el investigador de CS de la UNMSM sus datos de investigación	101
Tabla 22 Distribución de frecuencias: cantidad de datos de investigación que el investigador de CS de la UNMSM coloca disponibles a los demás	101
Tabla 23 Distribución de frecuencias: usuarios que tienen acceso a los datos de investigación de los investigadores de CS de la UNMSM	102
Tabla 24 Distribución de frecuencias: medios a través de los cuales se accede a los datos de investigación de los investigadores de CS de la UNMSM	103

Tabla 25 Distribución de frecuencias: utilización de datos generados por otros investigadores o instituciones, realidos por los investigadores de CS de la UNMSM..	104
Tabla 26 Distribución de frecuencias: reutilización de datos generados por otros investigadores o instituciones, realizados por los investigadores de CS de la UNMSM.....	105
Tabla 27 Distribución de frecuencias: tipo de licencia que aplican los investigadores de CS de la UNMSM	106
Tabla 28 Distribución de frecuencias: como registran sus datos de investigación los investigadores de CS de la UNMSM	107
Tabla 29 Distribución de frecuencias: tipo de formato que usa el investigador de CS de la UNMSM para almacenar sus datos de investigación	108
Tabla 30 Distribución de frecuencias: técnicas de recolección de datos que utiliza el investigador de CS de la UNMSM	109
Tabla 31 Distribución de frecuencias: tipo de datos de investigación que utilizan los investigadores de CS de la UNMSM	110
Tabla 32 Distribución de frecuencias: condiciones requeridas por los investigadores de CS de la UNMSM para enviar sus datos de investigación a un repositorio.....	111
Tabla 33 Distribución de frecuencias: espacio de almacenamiento que necesita los investigadores de CS de la UNMSM	112
Tabla 34 Distribución de frecuencias: tipo de almacenamiento actualmente utiliza el investigador de CS de la UNMSM	113
Tabla 35 Distribución de frecuencias: razones más importantes para los investigadores de CS de la UNMSM para que sus datos de investigación estén disponibles y preservados	115
Tabla 36 Distribución de frecuencias: opciones de apoyo que el investigador de CS de la UNMSM esperaría de su institución si implementará un repositorio	116
Tabla 37 Distribución de frecuencias: preferencia del responsable para hacerse cargo del archivo de datos de investigación de los investigadores de CS de la UNMSM.	118
Tabla 38 Distribución de frecuencias: instituciones donde los investigadores de CS de la UPC realizan sus trabajos de investigación.....	119
Tabla 39 Distribución de frecuencias: líneas de investigación a las que se dedican los investigadores de CS de la UPC	120
Tabla 40 Distribución de frecuencias: agencias que contribuyen con fondos a la investigación en UPC.....	122
Tabla 41 Distribución de frecuencias: uso que le da el investigador de CS de la UPC	123
Tabla 42 Distribución de frecuencias: tiempo que conserva el investigador de CS de la UPC sus datos de investigación	123
Tabla 43 Distribución de frecuencias: cantidad de datos de investigación que el investigador de CS de la UPC coloca disponibles a los demás	124
Tabla 44 Distribución de frecuencias: usuarios que tienen acceso a los datos de investigación de los investigadores de CS de la UPC.....	125

Tabla 45 Distribución de frecuencias: medios a través de los cuales se accede a los datos de investigación de los investigadores de CS de la UPC.....	127
Tabla 46 Distribución de frecuencias: utilización de datos generados por otros investigadores o instituciones, realizados por los investigadores de CS de la UPC	128
Tabla 47 Distribución de frecuencias: reutilización de datos generados por otros investigadores o instituciones, realizados por los investigadores de CS de la UPC	128
Tabla 48 Distribución de frecuencias: tipo de licencia que aplican los investigadores de CS de la UPC.....	129
Tabla 49 Distribución de frecuencias: sobre como registran sus datos de investigación los investigadores de CS de la UPC	130
Tabla 50 Distribución de frecuencias: tipo de formato que usa el investigador de CS de la UPC para almacenar sus datos de investigación	132
Tabla 51 Distribución de frecuencias: técnicas de recolección de datos que utiliza el investigador de CS de la UPC.....	133
Tabla 52 Distribución de frecuencias: tipo de datos de investigación que utilizan los investigadores de CS de la UPC	134
Tabla 53 Distribución de frecuencias: condiciones requeridas por los investigadores de CS de la UPC para enviar sus datos de investigación a un repositorio	135
Tabla 54 Distribución de frecuencias: espacio de almacenamiento que necesitan los investigadores de CS de la UPC	137
Tabla 55 Distribución de frecuencias: tipo de almacenamiento actualmente utiliza el investigador de CS de la UPC.....	138
Tabla 56 Distribución de frecuencias: razones más importantes para los investigadores de CS de la UPC para que sus datos de investigación estén disponibles y preservados.....	139
Tabla 57 Distribución de frecuencias: opciones de apoyo que el investigador de CS de la UPC esperaría de su institución si implementará un repositorio.....	140
Tabla 58 Distribución de frecuencias: preferencia del responsable para hacerse cargo del archivo de datos de investigación de los investigadores de CS de la UPC.....	142
Tabla 59 Distribución de frecuencias: instituciones donde los investigadores de la UNMSM - UPC realizan sus trabajos de investigación	143
Tabla 60 Distribución de frecuencias: líneas de investigación a las que se dedican los investigadores de CS de la UNMSM - UPC	144
Tabla 61 Distribución de frecuencias: agencias que contribuyen con fondos a la investigación en UNMSM - UPC	145
Tabla 62 Distribución de frecuencias: uso que le da el investigador de la UNMSM - UPC	146
Tabla 63 Distribución de frecuencias: tiempo que conserva el investigador de CS de la UNMSM - UPC sus datos de investigación	146
Tabla 64 Distribución de frecuencias: cantidad de datos de investigación que el investigador de CS de la UNMSM - UPC coloca disponibles a los demás	147
Tabla 65 Distribución de frecuencias: usuarios que tienen acceso a los datos de investigación de los investigadores de CS de la UNMSM - UPC	148

Tabla 66 Distribución de frecuencias: medios a través de los cuales se accede a los datos de investigación de los investigadores de CS de la UNMSM - UPC	149
Tabla 67 Distribución de frecuencias: utilización de datos generados por otros investigadores o instituciones, por los investigadores de CS de la UNMSM - UPC	150
Tabla 68 Distribución de frecuencias: reutilización de datos generados por otros investigadores o instituciones, por los investigadores de CS de la UNMSM - UPC	151
Tabla 69 Distribución de frecuencias: tipo de licencia que aplican los investigadores de CS de la UNMSM - UPC	152
Tabla 70 Distribución de frecuencias: sobre como registra sus datos de investigación los investigadores de CS de la UNMSM - UPC	153
Tabla 71 Distribución de frecuencias: tipo de formato que usa el investigador de CS de la UNMSM - UPC para almacenar sus datos de investigación	154
Tabla 72 Distribución de frecuencias: técnicas de recolección de datos que utiliza el investigador de CS de la UNMSM - UPC	155
Tabla 73 Distribución de frecuencias: tipo de datos de investigación que utilizan los investigadores de CS de la UNMSM - UPC	156
Tabla 74 Distribución de frecuencias: condiciones requeridas por los investigadores de CS de la UNMSM - UPC para enviar sus datos de investigación a un repositorio	157
Tabla 75 Distribución de frecuencias: espacio de almacenamiento que necesita los investigadores de CS de la UNMSM - UPC	159
Tabla 76 Distribución de frecuencias: tipo de almacenamiento actualmente utiliza el investigador de CS de la UNMSM - UPC	160
Tabla 77 Distribución de frecuencias: razones más importantes para los investigadores de CS de la UNMSM - UPC para que sus datos de investigación estén disponibles y preservados	161
Tabla 78 Distribución de frecuencias: opciones de apoyo que el investigador de CS de la UNMSM - UPC esperaría de su institución si implementará un repositorio ..	162
Tabla 79 Distribución de frecuencias: preferencia del responsable para hacerse cargo del archivo de datos de investigación de los investigadores de CS de la UNMSM - UPC	164

ÍNDICE DE ANEXOS

Anexo 1 Glosario	181
Anexo 2 Relación de Facultades de Medicina: Asociación Peruana de Facultades de Medicina (ASPEFAM)	185
Anexo 3 Documentos de trabajo, Diseño del Modelo propuesto	186

INTRODUCCIÓN

A finales del siglo XX, la comunicación científica sufrió un gran cambio debido a la crisis del sistema tradicional de comunicación científica, en el cual los precios de las revistas científicas subieron considerablemente frente a la disminución de los presupuestos de las instituciones para la adquisición de las mismas, debido a esta situación nace el movimiento Open Access o Acceso Abierto. La iniciativa del Acceso Abierto fue promovido en sus inicios por SPARC (Scholarly Publishing & Academic Resource Coalition) en 1997, que estableció una colaboración entre bibliotecas académicas y sociedades o instituciones científicas. Esta iniciativa se transformó en lo que actualmente se conoce como el movimiento del Acceso Abierto, con la Declaración de Budapest en el año 2002, momento en el cual alcanza su mayor desarrollo, ya que fue la primera Declaración en establecer una definición pública del acceso abierto como la “disponibilidad gratuita en la Internet Pública, para que cualquier usuario la pueda leer, descargar, copiar, distribuir, imprimir, con la posibilidad de buscar o enlazar todos los textos de estos artículos, recorrerlos para indexación exhaustiva, usarlos como datos para software, o utilizarlos para cualquier otro propósito legal, sin barreras financieras, legales o técnicas, distintas de la fundamental de ganar acceso a la propia Internet”. ([http:// www.soros.org/openaccess/translations/spanish-translation](http://www.soros.org/openaccess/translations/spanish-translation)). En esta Declaración se proponen también, las estrategias para hacer realidad el acceso abierto a través de los repositorios llamado “acceso abierto verde” y de las revistas llamado “acceso abierto dorado”. Al año siguiente, 2003, se afianzó esta iniciativa con la Declaración de Berlín, la cual establece dos condiciones para el acceso abierto, siendo la segunda condición el “que una versión completa del trabajo y todos los materiales complementarios, que incluya una copia del permiso del autor, en un formato electrónico estándar, sea depositado y publicado en un repositorio electrónico que utilice estándares técnicos aceptables (tales como las definiciones del Acceso Abierto)”, (http://openaccess.mpg.de/67627/Berlin_sp.pdf). Ese mismo año la Declaración de Bethesda sobre Publicación de Acceso Abierto define a la investigación científica y sus objetivos como “un proceso interdependiente donde cada

experimento es informado por el resultado de otros. Así mismo declara que las publicaciones electrónicas de resultados de investigación ofrecen la oportunidad y la obligación de compartir resultados de investigación, ideas y descubrimientos libremente con la comunidad científica y el público”. (<http://www.earlham.edu/~peters/fos/bethesda.htm>). En el Perú la definición de Acceso Abierto se encuentra en el 2º, acápite 2.1 del Reglamento de la Ley 30035, Ley que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto que define el Acceso Abierto como “uso lícito que confiere un titular de derechos de propiedad intelectual a cualquier persona, para que pueda acceder de manera inmediata y gratuita a una obra, datos procesados o estadísticas de monitoreo, sin necesidad de registro, suscripción, ni pago, estando autorizada a leerla, descargarla, reproducirla, distribuirla, imprimirla, buscarla y enlazar textos completos”.

El presente estudio busca descubrir las prácticas que utilizan los investigadores de la Facultad de Ciencias de la Salud de la Universidad Nacional Mayor de San Marcos (UNMSM) y la Facultad de Ciencias de la Salud de la Universidad Peruana de Ciencias Aplicadas (UPC), para la gestión de sus datos de investigación y a partir de esta ello, elaborar una propuesta para optimizarlas. En ese sentido, el propósito de este estudio es diseñar un modelo para la gestión de datos de investigación basados en el ciclo de vida de los datos que contenga como elemento crítico la difusión de los beneficios de compartir y reutilizar los datos de investigación por los investigadores de ambas universidades en el área de las Ciencias de la Salud.

Este estudio consta de los siguientes seis capítulos:

En el primer capítulo se desarrolla en forma general la situación problemática, la formulación del problema, la justificación de la investigación y el objetivo general que es: elaborar un modelo del ciclo de vida para la gestión de datos de investigación en áreas de ciencias de la Salud de universidades.

En el segundo capítulo se describe la revisión bibliográfica sobre las teorías que fundamentan el marco teórico de este estudio. Se mencionan los antecedentes,

basados en las experiencias de otras investigaciones sobre datos de investigación; gestión de datos en repositorios de datos; e implementación de repositorios de datos en algunas universidades a nivel internacional. Del mismo modo, se definen los conceptos de datos de investigación; los elementos que forman parte de la gestión de los datos; la gestión de datos basados en el ciclo de vida de los mismos; la importancia de la preservación de los datos; los lineamientos y políticas de gestión de datos; estándares sobre el uso de metadatos; y software para la gestión de los datos a nivel nacional e internacional. Finalmente describe el campo de investigación en Ciencias de la Salud; y explica sobre las áreas que conforman cada una de las Facultades de Ciencias de la Salud de las universidades objetos de este estudio: Universidad Nacional Mayor de San Marcos (UNMSM) Universidad Peruana de Ciencias Aplicadas (UPC).

En el tercer capítulo se expone la propuesta de un modelo para la gestión de los datos de investigación, en base al ciclo de vida de los datos, con la finalidad de promover más investigaciones y por lo tanto generar nuevo conocimiento, minimizando esfuerzos y recursos, bajo el enfoque del acceso abierto, así también como los beneficios que aporta la propuesta a esta investigación.

En el cuarto capítulo se hace referencia a la metodología que se ha utilizado en este estudio, mostrando el tipo y diseño de la investigación; las variables de estudio; indicadores; la unidad de análisis; la población; la muestra; la técnica de recolección de datos; el instrumento; y los ítems.

En el quinto capítulo se presenta el cuestionario, se muestra el análisis e interpretación de los resultados de la Universidad Nacional Mayor de San Marcos (UNMSM) y de la Universidad Peruana de Ciencias Aplicadas (UPC), se expone el análisis comparativo e interpretación de los resultados de la encuesta y se presentan las conclusiones de la encuesta realizada en ambas universidades.

En el sexto capítulo se expone las conclusiones generales de la investigación respecto al objetivo general y a los objetivos específicos.

CAPÍTULO I: SITUACIÓN PROBLEMÁTICA

1.1. Situación problemática.

Según Torres- Salinas, et.al. (2011) en su artículo *Compartir los datos de investigación en ciencia: introducción al data sharing* dice que “durante la última década la comunidad científica está debatiendo intensamente la necesidad de compartir en acceso abierto los datos fruto de las investigaciones para que éstos puedan ser reutilizados con diferentes propósitos por el resto de investigadores, una tendencia que se conoce en inglés como data sharing”. (p. 174). En este sentido podemos mencionar que, si bien desde hace muchos años se han establecido diversas formas de compartir información, ya sea a través de bancos de datos o a través de repositorios institucionales, es a partir de este siglo que se ha empezado a hablar sobre la iniciativa de datos abiertos de investigación, es decir datos que puedan ser accesibles, compartidos y reutilizados por los investigadores. El compartir información es algo que ya se viene realizando a nivel mundial y a través de repositorios de datos con muy buenos resultados, como es el caso del artículo publicado por Barrios, E., et al (2011), Implementación de un repositorio de datos científicos usando DSpace en el cual se demuestra que el repositorio LAGODATOS, almacena los datos adquiridos de los países miembros de LAGO (Argentina, Bolivia, Colombia, México, Perú y Venezuela) y se encuentran disponibles para la comunidad.

De acuerdo a lo expresado por Hernández-Pérez, et.al. (2013), en su obra *Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios*, “los problemas con el acceso a los datos no son tanto de infraestructura tecnológica, pues en algunos casos se colocan en la plataforma del repositorio de la institución, sino que los problemas se refieren a los siguientes aspectos: actitud recelosa de los investigadores para compartir sus datos. Hay que concientizar a los investigadores sobre la importancia de compartir datos. Actualmente este punto se está solucionando con las agencias financiadoras quienes están obligando a los investigadores a colocar sus datos de investigación; la falta de planificación de cómo gestionar estos datos. La gestión de los datos requerirá un reciclaje tanto de los profesionales, como de los planes de

formación para poder gestionar correctamente un repositorio de datos, debido a la gran cantidad de datos, sobre todo numéricos que tendrán los repositorios; la falta de formación de investigadores, gestores de información y bibliotecarios para gestionar el ciclo de vida de datos científicos y; el almacenamiento de los datos que se encuentran dispersos, debido a que aún no existe conciencia de la necesidad para poder guardar y preservarlos”. Hernández-Pérez, (2013: p.263.). De acuerdo a los resultados de las encuestas seleccionadas para realizar este caso de estudio en relación a las actitudes y prácticas de los investigadores en el uso y reutilización de datos de biomedicina realizado por Federer, L. y publicado PlosOne (2011), y por el CNR (National Research Council), los investigadores si bien comparten sus datos, no los comparten en su totalidad, siendo uno de los problemas más importantes la garantía de que los datos se vuelvan a utilizar de forma responsable.

Desde estas perspectivas, el uso compartido de los datos permite verificar los resultados de la investigación y ampliar la verificación a partir de los resultados anteriores, así mismo, ayuda a realizar un análisis más eficaz de los datos de investigación. Aunque actualmente, existen muchos argumentos a favor de compartir datos hay muchos investigadores que si bien los comparten, lo hacen a través de vías informales, es decir: por petición de un colega de su propia disciplina; por recelo para compartir sus datos debido a que no tienen el tiempo suficiente para gestionarlos; por su preocupación de futuras oportunidades de publicación; por el deseo de retener los derechos exclusivos sobre los datos debido al tiempo y esfuerzo que les ha tomado elaborarlos y por el temor de que otros puedan cuestionar o malinterpretar los resultados o conclusiones de la recopilación de datos. A su vez, estos datos se encuentran dispersos y guardados por los investigadores en diversos dispositivos, en sus bitácoras, en sus computadoras personales u otros lugares que no son accesibles para otros investigadores, ocasionando en algunos casos la pérdida de datos valiosos y de conocimiento para el mundo científico.

Por otro lado, existen lineamientos y políticas a nivel mundial, planteados por los gobiernos e instituciones académicas, sobre el tratamiento de estos datos, porque permite reducir los costos, aumentar la transparencia y verificar los resultados de la

investigación como es el caso de la Unión Europea, Estados Unidos, Reino Unido, Nueva Zelanda, Canadá y en Latinoamérica como es el caso de México y Argentina. En el Perú, también existen lineamientos y políticas a nivel gobierno, como es el caso de la promulgación de la Ley 30035, Ley que regula el Repositorio Nacional Digital de Ciencias, Tecnología e Innovación de Acceso Abierto, la tendencia es que los datos abiertos de investigación sean accesibles, compartidos y reutilizados por los investigadores, situación que favorecería el desarrollo de la investigación en el Perú. Sin embargo, en las universidades aún no se han dado lineamientos específicos de cómo gestionar los datos de investigación.

1.2. Formulación del problema principal

En el campo de las Ciencias de la Salud, la generación de conocimiento pasa por el incremento del número de investigaciones que actualmente no se realiza, debido a que los investigadores no comparten sus datos de investigación ni reutilizan los datos de investigación de otros investigadores, por lo tanto, se presenta la necesidad de plantear, en el marco en el que se debe dar este proceso, un modelo de gestión de datos de investigación, que no hemos podido encontrar fácilmente, que se ajuste a las realidades de las universidades públicas y privadas en el Perú.

La investigación en el área de ciencias de la salud de las universidades, no progresa debido a que no existe el hábito o costumbre de compartir o reutilizar los datos de investigación entre los investigadores. Así mismo no se encuentra un modelo de gestión de datos de investigación que generan los investigadores, del mismo modo no está sistematizada la creación de datos de investigación, la evaluación y selección de los datos a gestionar, la preservación o mantenimiento de los datos, la reutilización de los datos y el almacenamiento de los mismos. En el ámbito de la investigación universitaria en la UNMSM y en la UPC en las áreas de ciencias de la Salud la tendencia actual es dar mayor énfasis a la investigación, para lo cual es fundamental el gestionar los datos de investigación que puedan ser compartidos y reutilizados entre los investigadores. Por tal razón, el número de investigaciones en el Perú es mucho menor que en otros países de la región como Argentina, Colombia, México entre otros.

Vista esta situación consideramos necesario analizar los modelos de ciclo de vida de los datos de investigación, y a partir de ello desarrollar un modelo que se ajuste a la realidad de los investigadores del área de Ciencias de la Salud que contribuya a compartir los datos de investigación, aumentar el número de investigaciones en las universidades y a su vez mejorar la calidad de las mismas.

1.2.1. Problemas específicos

1. Los investigadores no usan los datos de investigación generados por otros investigadores.
2. La falta de organización de los datos de investigación no hace posible la gestión de los datos de investigación
3. Se dificulta el proceso de almacenamiento de los datos de investigación por qué no se gestionan adecuadamente los datos de investigación.

1.3. Justificación teórica

En la actual sociedad del conocimiento, la generación de datos e información se ha venido incrementando de manera exponencial, frente a esta situación ha sido necesario plantear métodos y establecer políticas y lineamientos que permitan gestionar la producción de la información de manera eficaz y en particular la gestión de los datos abiertos. Uno de los métodos planteados es realizar la gestión de datos de investigación basados en el ciclo de vida de los datos,

De acuerdo a lo señalado en Hernández-Pérez y García-Moreno (2013, “la definición de “abierto”, se entiende dentro del contexto que aprueba el Open Definition Advisory Council, una organización con instituciones y representantes personales destacados del movimiento de acceso abierto en donde se señalan 11 condiciones para qué una obra o unos datos sean considerados abiertos: acceso (disponible integralmente, a un coste razonable y de forma que pueda ser modificable), redistribución, reutilización, ausencia de restricciones tecnológicas, reconocimiento,

integridad, sin discriminación de personas o grupos, sin discriminación de ámbitos de trabajo, distribución de la licencia”. (p. 262).

En las Directrices sobre la aplicación del acceso abierto a las publicaciones científicas y a los datos de investigación en proyectos apoyados por el Consejo Europeo de Investigación en el marco de Horizonte 2020 European Research Council (2016) señala que “el Consejo Europeo de Investigación (ERC) apoya el principio básico de acceso abierto a los datos de investigación. Por lo tanto, recomienda a todos sus investigadores financiados que sigan las mejores prácticas al retener los archivos de todos los datos de investigación que han producido y utilizado durante su trabajo y que estén dispuestos a compartir estos datos con otros investigadores siempre que no estén vinculados por restricciones de derechos de autor, requisitos de confidencialidad o cláusulas contractuales”.

Según Pérez-Hernández y García-Moreno (2013) “Las tecnologías de la información están permitiendo recopilar grandes cantidades de datos: datos personales; datos sobre el contexto en que vivimos; datos sobre objetos y productos; datos sobre los procesos, sin embargo, no todos los datos que se encuentren en la web son datos abiertos, para que presenten esta característica es necesario que posean algún tipo de licencia legal para poderlos utilizar, reutilizar y redistribuir”. (2013: p.259). Desde este enfoque, las tecnologías de la información son herramientas que nos pueden permitir recopilar datos de investigación que sirvan para crear o generar nuevo conocimiento. Para que estos datos puedan ser utilizados, reutilizados y redistribuidos con fines académicos y de investigación es necesario comprender las consideraciones legales y económicas que faciliten el acceso abierto a estos datos.

En los últimos años, en América Latina a través de la iniciativa “Alianza para el Gobierno Abierto (AGA), apoyada por la OCDE, se promueve mejorar los niveles de transparencia de los datos que maneja la Administración Pública a fin de rendir cuentas a los ciudadanos. En el Perú, el desarrollo de repositorios institucionales y de datos se encuentra en su etapa emergente a raíz de la Ley 30035 y su Reglamento, este último define a los datos procesados como información obtenida después de haber realizado

una serie de operaciones sobre un conjunto de datos. A partir de la promulgación de la Ley de Transparencia y Acceso a la Información Pública. Ley 27806 se promueve la publicación de datos abiertos de las universidades públicas en el Portal de Transparencia Estándar (PTE), sin embargo, los datos publicados corresponden más a datos generales y administrativos, pero aún no está desarrollado el promover la publicación de datos abiertos de investigación en instituciones educativas, privadas y públicas.

Valentín (2013) describe la gestión de los datos, basado en el ciclo de vida de los datos. En esta tesis hace mención al modelo que presenta el Digital Curation Center y que contiene una visión general de las etapas requeridas para realizar una gestión activa de los datos de investigación, permite también definir roles y responsabilidades en la gestión de los datos de acuerdo con el ciclo de vida de los mismos.

En la tesis revisada de Pérez M. (2012) titulada *Una Aproximación al data curation y el rol del bibliotecario en su implantación: el caso de la Universitat Oberta de Catalunya*, señala “que los riesgos de la falta de gestión de datos son la creación de datos de forma indiscriminada y sin normalización ni control institucional, lo que conlleva un alto riesgo de pérdida de información a largo plazo, de escasa reutilización e interoperabilidad de los mismos”. Señala además (Pryor, 2012) “que para una correcta gestión de datos muchos organismos han propuesto una gestión de datos basándose en su ciclo de vida, lo cual permite asegurar que se desarrollen políticas y procesos, se definan roles y responsabilidades y se cree un marco tecnológico adecuado para su almacenamiento, es decir estándares y tecnologías a desarrollarse”.

En ese sentido, los beneficios del acceso abierto a los datos de investigación pasarán por facilitar a los investigadores acceder a información valiosa sin incurrir en el uso de recursos escasos como son el tiempo y el dinero. Así mismo, la trascendencia de contar con el acceso abierto a los datos de investigación implica la mejora de la calidad de la información, la reducción en el tiempo de investigación, la visualización del investigador, el estímulo para que muchos investigadores utilicen repositorios especializados en las universidades y la adopción de políticas nacionales. Por tal motivo, se presenta la necesidad de realizar este estudio a fin de conocer la percepción

de los investigadores sobre la necesidad de usar, compartir y reutilizar los datos de investigación.

1.4. Justificación práctica

Algunas universidades en el mundo han desarrollado repositorios institucionales que almacenan los resultados de investigación de sus miembros, sin embargo, en el Perú, las universidades, aún no han puesto en marcha modelos apropiados para la gestión de los datos que se producen como resultado de una investigación, y como consecuencia de ello, se pierde la posibilidad para su publicación en alguna revista científica, y su uso como punto de partida para otra investigación, es nula.

Valentín (2013) sostiene que se debe tener en cuenta el ciclo de vida de los datos, es decir desde su creación y almacenamiento hasta que se determine su obsolescencia y sean desechados pasando por un proceso de selección, conservando solamente aquellos que han servido como resultado para la investigación. Describe que el plan de gestión de datos es la parte más importante de este proceso y debe considerar el ciclo de vida de los datos.

Por otro lado, tal como los sostienen Torres-Salinas, Robinson-García y Cabezas-Clavijo (2011) respecto a la gestión de datos, el fin es: "... minimizar recursos y maximizar esfuerzos, así como también obtener un mayor rendimiento del dinero invertido en ciencia, dado que se evitaría la duplicación de proyectos destinados a obtener resultados similares y se preservarían estos datos a través de una infraestructura adecuada. Todo ello, permitirá de algún modo mostrar la importancia de implementar un repositorio de datos a fin de compartir información de una determinada disciplina y evitar los canales informales que se utilizan actualmente para poder compartir información como por ejemplo respondiendo a solicitudes concretas de otros profesionales de la misma disciplina sin llegar a ser compartidos de manera pública; y de forma descentralizada". (2011: p.174). De acuerdo con los autores, podemos exponer las razones prácticas que justifican un estudio sobre la gestión de los datos de investigación coleccionados en un repositorio institucional:

- Gestionar los datos de investigación basados en el ciclo de vida de los datos con el fin de permitir el acceso abierto que faciliten la visibilidad a los datos de investigación, para que sean compartidos de manera pública y de forma descentralizada.
- Concentrar los datos de investigación por especialidad en un repositorio, permitirá manejar mejor, en términos prácticos, los recursos financieros destinados para este tipo de proyectos.
- Conjuguar esfuerzos para la preservación idónea de los datos de investigación en tener una infraestructura adecuada.
- Agrupar los datos de investigación por especialidad, los investigadores en ese campo especializado podrían compartir, utilizar y reutilizar estos datos, de manera óptima;
- Preservar la integridad de la investigación, es decir, evitar la no contaminación de la información, a través de lineamientos y políticas, que cumplan con asegurar que los datos estén disponibles y puedan ser reutilizados de manera continua y sin restricción alguna.

1.5. Objetivos de la investigación

1.5.1. Objetivo general

Elaborar un modelo del ciclo de vida para la gestión de datos de investigación en áreas de ciencias de la Salud de universidades con la finalidad de generar nuevo conocimiento.

1.5.2. Objetivos específicos

- Determinar el uso de los datos de investigación generados por los investigadores, para que estén disponibles y puedan ser accedidos fácilmente por otros investigadores.

- Determinar la organización de los datos de investigación para el acceso a los datos generados por los investigadores en las universidades con el propósito de que los datos de investigación sean reutilizados.
- Determinar el proceso de almacenamiento de los datos generados por los investigadores en las universidades para que estos sean preservados.

1.6. Variables de estudio

1.6.1. Variable independiente

Gestión de datos de investigación

1.6.2. Variables dependientes

Acceso a los datos de investigación:

- Uso de los datos de investigación
- Organización de los datos de investigación
- Almacenamiento de los datos de investigación

1.7. Operacionalización de variables

1.7.1. Variable dependiente 1

Uso de los datos de investigación

Indicadores:

- Conservación de datos de investigación
- Duración de la conservación de los datos de investigación
- Disponibilidad de datos de investigación
- Acceso a entidades
- Medios de acceso a los datos de investigación por otros investigadores
- Datos generados utilizados por otros investigadores
- Reutilización de los datos de investigación

1.7.2. Variable dependiente 2

Organización de los datos

Indicadores:

- Software estándar
- Uso de formatos de almacenamiento
- Técnicas de recolección de datos
- Tipo de datos de investigación

1.7.3. Variable dependiente 3

Almacenamiento de datos de investigación

Indicadores:

- Condiciones de envío de datos de investigación
- Espacio de almacenamiento utilizado por un investigador
- Tipo de almacenamiento utilizado por un investigador
- Razones de disponibilidad y preservación
- Opciones de apoyo respecto a la implementación del repositorio de datos de investigación
- Responsable de la gestión de datos de investigación

CAPÍTULO II: MARCO TEÓRICO

2.1. Introducción

El marco teórico de la presente investigación, tiene por finalidad exponer los aspectos más importantes de la revisión bibliográfica realizada con este fin. En ese sentido, se han revisado conceptos vinculados a todas aquellas teorías que fundamentan: la gestión de datos de investigadores en las facultades de ciencias de la salud de las universidades peruanas, Universidad Nacional Mayor de San Marcos (UNMSM), y Universidad Peruana de Ciencias Aplicadas (UPC); repositorios de datos en universidades internacionales y nacionales; datos de investigación; conceptos y modelos sobre el ciclo de vida de datos de investigación; preservación de los datos; lineamientos y políticas de gestión de datos, acceso a datos abiertos promovidos por los gobiernos y entidades involucradas en la gestión de datos abiertos a través de la implementación de portales de datos de investigación; en estándares para el uso de metadatos; y softwares para la gestión de datos. Este marco contribuirá a que todas aquellas personas interesadas en el tema de gestión de datos de investigación, puedan entender los conceptos esenciales en este campo.

2.2. Antecedentes de la investigación

El artículo en idioma inglés, publicado por Katherine Mc Neil, (Mc-Neil 2007), *Interoperability between institutional and data repositories: a pilot project at MIT*, describe que en el Instituto Tecnológico de Massachusetts (MIT), los bibliotecarios desempeñan un papel clave en la promoción y facilitación del uso de los repositorios institucionales (Bailey 2005) y su interoperabilidad con un repositorio de datos. Estas actividades proporcionan nuevas oportunidades a los bibliotecarios quienes se convierten en socios de los profesores, apoyándolos en la publicación de sus contenidos, lo cual puede enriquecer sus relaciones y aumentar la relevancia de la biblioteca (Buehler y Boateng 2005; Bell, Foster y Gibbons 2005).

El autor señala que los miembros de la Facultad de Ciencias Sociales del MIT tienen tres opciones donde pueden depositar los datos que han producido: DSpace, software del repositorio institucional que preserva, no solamente los conjuntos de datos producidos por los profesores, sino que también preserva cualquier material (documentos, imágenes etc.) que trabajan. El software Dataverse (DVN), desarrollado en Harvard y personalizado para el MIT (King 2007), en el cual los miembros de la facultad pueden depositar sus datos y el software del Consorcio Interuniversitario para Política Social e Investigación (ICPSR). Cada uno de estos sistemas tiene sus propias ventajas y desafíos según las necesidades de cada uno de los miembros de los departamentos de economía, historia y ciencias políticas el depósito de sus datos. El bibliotecario ha desempeñado un papel vital en considerar un proyecto para desarrollar un prototipo de sistema que permita la interoperabilidad de estos tres sistemas en relación al intercambio de contenidos y metadatos. Este sistema se denomina PROMESA. En este proyecto se seleccionaron los sistemas DSpace y DVN porque son de cosecha propia y el MIT tiene participación en el desarrollo del sistema. Finalmente, ahora los metadatos de DVN están disponibles como un archivo adicional asociado con la partida de DSpace.

Barrios, Torrén, Núñez y Torres (2011), señalan en su artículo *Implementación de un repositorio de datos científicos usando DSpace* “respecto a los datos e-investigación, que “los datos pueden ser el comienzo (o corroboración) de las ideas y, consecuentemente, deberían ser de libre acceso para que con su uso y reutilización se puedan seguir la cadena de producción del conocimiento: datos-información-conocimiento-información-datos (p.102). Observamos que los datos para la generación y desarrollo del conocimiento son muy importantes por cuanto representan ideas que pueden ser utilizadas y reutilizadas para la generación del conocimiento. Barrios et al. nos describen que la comunidad científica que más se interesó en lo relacionado al modelo de e-investigación es la de Astrofísica” (p.103), Barrios et al. continúan explicando que “el proyecto LAGO es una colaboración entre 40 investigadores de 15 instituciones de América Latina” (p.104), lo que constituye un ejemplo de colaboración que investigadores de distintas instituciones se pueden juntar para desarrollar proyectos de investigación en conjunto. Barrios et al. también nos describen la Red de Repositorios LAGO Datos que “consiste en una infraestructura

diseñada para proveer a los investigadores de la colaboración LAGO de una herramienta para preservar, catalogar y difundir los datos (registrados y simulados) generados en cada una de las instalaciones, así como difundir publicaciones generadas por cada uno de los grupos de trabajo LAGO” (p.104). Al respecto podríamos decir que la implementación de un repositorio de datos o de una red de repositorios requiere de los recursos necesarios para proveer a los investigadores de las herramientas, la infraestructura y las instalaciones apropiadas que contribuirán a obtener buenos resultados.

La tesis de Pérez (2012) titulada *Una aproximación al data curation y el rol del bibliotecario en su implantación: el caso de la Universitat Oberta de Catalunya*, trata sobre la gestión, recuperación y preservación de los datos en dicha Universidad, así como también el papel que desempeña la biblioteca en este proceso. Establece un análisis sobre el significado de *data curation*, y el cambio de rol del bibliotecario como apoyo al investigador en todo el proceso de la gestión de los datos. Finalmente presenta un informe diagnóstico sobre la Universitat Oberta de Catalunya en relación a la gestión de los datos, proponiendo un conjunto de medidas y recomendaciones para mejorar la situación de los datos de investigación y el data curation. Según esta perspectiva se debe agregar al perfil de los bibliotecarios este nuevo rol. El bibliotecario o gestor de la información debe estar al tanto de las necesidades y requerimientos de los investigadores para la gestión de los datos, es decir desde la creación de los mismos hasta su almacenamiento en un repositorio de datos y su posterior reutilización. Así mismo, debe seleccionar con el investigador los datos que van a ser de corto o largo plazo y los que van a ser desechados, teniendo en cuenta aquellos datos que no cumplan con el fin del resultado de la investigación requerida.

De la revisión de la tesis de Valentín (2013) titulada *Gestión de datos de investigación*, de la Universidad Politécnica de Valencia, rescatamos el objetivo principal, que es dar a conocer, qué son los datos de investigación y cómo se gestionan. Describe definiciones proporcionadas por diferentes autores e instituciones tales como: la Organización para la Cooperación Económica y el Desarrollo Económicos (OCDE); National Science Board (NSBD); y National Institutes of Health (NIH); entre

otras; Así mismo, presenta una visión general del movimiento Open Access (OA). Hace énfasis en la importancia de la gestión y preservación de los datos, debido al aumento del volumen de datos valiosos, haciendo hincapié en dos proyectos internacionales sobre la gestión y preservación de datos de investigación como son: Digital Curation Center (DCC); y Australian National Data Service (ANDS). Valentín considera que es indispensable para la comunidad científica contar con una infraestructura para la gestión de los datos a fin de asegurar su buen almacenamiento, facilitar el acceso a los mismos y su reutilización. Finalmente, describe una serie de proyectos donde se está desarrollando esta iniciativa.

De la revisión de las tesis en idioma inglés tenemos:

Balachandra, (2003), señala en su tesis *Data Repository for Autonomous guided vehicle using ASP.NET*, que el uso de la tecnología.NET sirve para la gestión y desarrollo de un repositorio de datos para el equipo del Centro de Robótica de la Universidad de California, que sirva para almacenar, recuperar y compartir sus datos. Se trata de una experimentación para aprender y adoptar esta tecnología mediante el intercambio de información con métodos del equipo del Centro de Robótica de la Universidad con el fin de poder ubicar la información en forma remota. El repositorio de datos se construye utilizando ASP.Net como el marco y el VB.Net como el lenguaje de codificación, por su fácil operatividad y escalabilidad. El objetivo principal es minimizar una acumulación en el flujo de trabajo debido a la falta de información. Cada capítulo de la tesis describe diferentes componentes para la implementación del repositorio tales como: el diseño de la interface; la construcción de la base de datos; los métodos; los procedimientos desarrollados en ASP.NET; la implementación del repositorio de datos en ASP.NET; y; finalmente da recomendaciones para futuras mejoras. Así mismo, señala que el repositorio de datos es valioso en términos de centralización de la información y es una herramienta muy útil a la hora de compartir la información sin ningún retardo de tiempo. Por otro lado, indica que los repositorios de datos necesitan ser actualizados periódicamente para estar al día y que es importante tener cuidado para evitar la redundancia de los datos en todo momento y que los datos sean solamente manipulados por el administrador o por el gestor de bases de datos.

Howard (2005) señala en su tesis titulada *BioDig: Architecture for Integrating Heterogeneous Biological Data Repositories Using Ontology* la necesidad de integrar la información de los repositorios de datos biológicos dispares, que se encuentran en distintos sitios en la Word Wide Web, mediante la implementación de un sistema basado en ontologías que podrían facilitar la integración de datos. Debido a que cada uno de los repositorios de datos contienen un problema o tema diferente, los investigadores y científicos deben acceder a diferentes interfaces, ya que estas son únicas y personalizadas. Esto se agrava debido a la diversidad en los datos y la heterogeneidad de su representación en internet, lo que ha ocasionado la necesidad de realizar búsquedas a través de referencias cruzadas mediante hipervínculos, siendo un problema que no se ha podido solucionar fácilmente. Por otro lado, los formatos de datos son a menudo incompatibles entre fuentes de datos diferentes ya que cada fuente de datos funciona de manera autónoma en la actualización de su información, lo que implica que los datos podrían convertirse con facilidad en inconsistentes y en múltiples repositorios de datos. Esta tesis se centra en el problema de la heterogeneidad de representación, es decir, en cómo se representan los datos en bases de datos y proporciona una solución que tiene como objetivo apoyar un sistema de consulta que sirve para integrar datos de múltiples bases de datos heterogéneas. La solución consiste en una arquitectura flexible que utiliza ontologías (propiedades similares), para facilitar la automatización e integración de datos de múltiples bases de datos heterogéneas. Así mismo, la tesis aborda el problema en varios niveles: la interfaz del usuario; el procesador de consultas; los resultados del agregador; la capa de la comunicación; y la capa de datos. La contribución innovadora de la tesis es la incorporación de las ontologías en el sistema, cuya finalidad es facilitar el procesamiento de consultas mediante la resolución de conflictos semánticos durante la integración del esquema y la búsqueda de información relevante. El sistema también aborda el problema de los identificadores únicos globales para facilitar el intercambio y acceso a la información mediante la promoción del uso de LSID. (Life Science Identifier).

Albrecht (2010) en su tesis titulada *Design of a Data Repository for a long-running physics experiment*, trata sobre la implementación de un repositorio de datos

científicos de física diseñados en un software libre. Los datos son recogidos del laboratorio de datos GRAND, un proyecto realizado en la Universidad Notre Dame, denominado *Gamma Ray Astrophysics experiment at Notre Dame GRAND*. Este proyecto fue diseñado para estudiar las fuentes de radiación de rayos gamma y ahora se utiliza para estudiar las emisiones solares y atmosféricas. El objetivo de la implementación del repositorio es proporcionar un sistema de almacenamiento y análisis fiable y escalable para los datos, diseñar una interfaz amigable para poder acceder a los datos con el fin de identificar rápidamente las áreas de interés para su recuperación y un estudio adicional para producir mejores resultados. Es importante considerar que un repositorio de datos debe ser escalable a fin de que exista interoperabilidad entre los diferentes sistemas de almacenamiento sin perder ningún dato ni interrumpir temporalmente el acceso del usuario a los datos, evitando que la capacidad de almacenamiento del repositorio no sea suficiente, debido a que se genera gran cantidad de datos y por lo tanto gran cantidad de gigabytes. Además de esta dificultad, existen otras limitaciones en el análisis, debido a la gran cantidad de datos; en la identificación de los archivos que son necesarios para calcular los datos que se deben incluir en el repositorio; y en el gran desafío que es la indexación de los archivos. El sistema utilizado para el repositorio debe ser de fácil acceso a través de muchas plataformas, con la finalidad de que los datos puedan ser compartidos entre universidades y otras instituciones.

Rogova, (2010), en su tesis titulada *Treatment of imprecision in data repositories with the aid of KNOLAP*, señala que al acceder a los datos hay que tener en cuenta que estos pueden ser imprecisos. Sugiere que esta imprecisión se puede dar debido a que estos pueden estar incompletos, y ser poco fiables o ambiguos debido a los errores de medición, además de que también pueden estar basados en “*default*” y los valores por defecto pueden tener excepciones. Sin embargo, cuando se tratan de datos científicos, estos deben ser organizados en torno a varios ejes de análisis y deben estar basados en jerarquías rígidas o flexibles. El objetivo de esta investigación es que todos los resultados de la investigación de la Universidad de Westminster estén a disposición del público a través de un repositorio de datos. Los datos además de estar organizados, integrados y almacenados, de acuerdo con varias dimensiones dependiendo de los intereses de los diferentes grupos de usuarios. El desafío más

importante para reducir la imprecisión de los datos es la clasificación de los mismos. Rogova, presenta varias soluciones para tratar la imprecisión de los datos, sin embargo la de mayor aceptación de esta investigación es KNOLAP, software ampliado del OLAP, que permite agilizar la consulta en línea de grandes cantidades de datos. Dentro de las otras soluciones describe el H-IFS y el IF Cube, referidas al manejo de las jerarquías.

Por todo lo expuesto podemos concluir que un punto importante a tomar en cuenta al momento de diseñar un repositorio de datos es el intercambio de información.

2.3. Revisión Bibliográfica

2.3.1. Gestión de datos de investigación

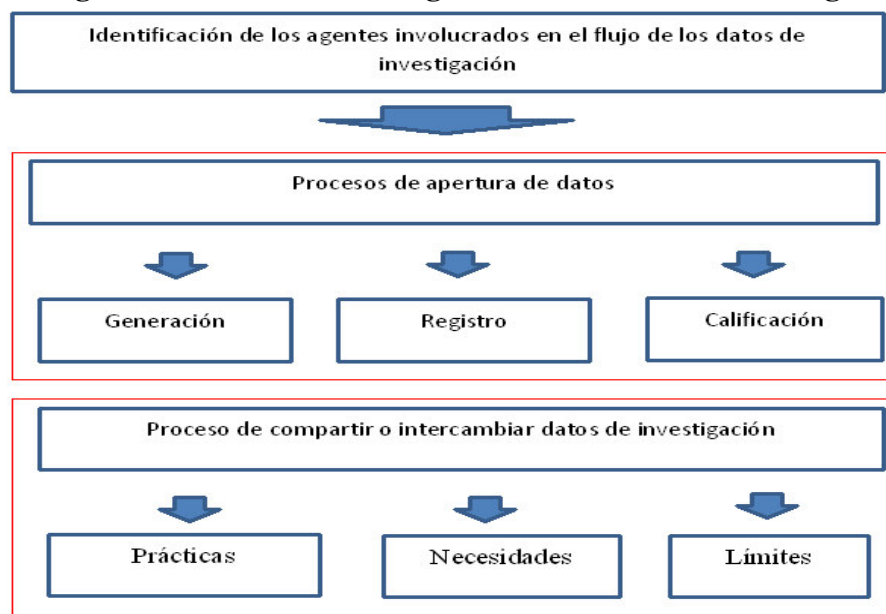
Gómez, Méndez y Hernández-Pérez (2016), en el artículo *Datos y metadatos de investigación en ciencias sociales y humanidades: una aproximación desde los repositorios temáticos de datos*, señalan “que compartir datos de investigación se ha convertido en una práctica habitual en disciplinas en las que existe una cultura científica muy colaborativa, como la física, la astronomía o la genética. A esta cultura disciplinaria se une además el hecho de que las instituciones públicas que financian la investigación han comenzado a exigir a los investigadores que hagan públicos sus resultados no sólo en forma de publicaciones, sino también abriendo los datos subyacentes utilizados” (p.546). Sostienen además que a pesar que esta tendencia de compartir los datos varía de una disciplina a otra, en general son muchas las motivaciones y ventajas que trascienden a los mandatos o tendencias a la hora de compartir los datos (Lyon, 2016) como por ejemplo: aumenta la posibilidad de incrementar el impacto y la visibilidad de la investigación; potencia la reproducibilidad de la ciencia; ahorra costos a la hora de crear datos; fomenta la colaboración; y contribuye a aumentar la credibilidad en el sistema” (p. 546,547).

Millán Gonzáles, Saorían, Ferrer-Sapena, Aleixander-Benavent y Peset (2013), en su artículo *Gestión de datos de investigación: infraestructura para su difusión*,

señalan que en relación con la definición sobre datos abiertos de investigación, la más aceptada es la que se adecúa a la adoptada por las entidades como National Institute of Health (NIH) de los Estados Unidos y la OECD que consideran “datos de investigación todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza. (p.416). A partir de la definición expuesta por la NIH, planteamos el desarrollo de este estudio. Estos mismos autores señalan que, en todo proceso para la gestión de los datos, es necesario que se tengan las personas involucradas y su participación:

1. “Quiénes son los agentes involucrados en el flujo de los datos de investigación;
2. Cuáles son sus prácticas, necesidades y límites a la hora de compartir sus datos; y
3. Qué datos se están generando, registrándolos y calificando especialmente su grado de apertura” (Millán Gonzáles 2et al., 2013, p. 421). (Ver gráfico 1).

Gráfico 1 Agentes involucrados en la gestión de los datos de investigación



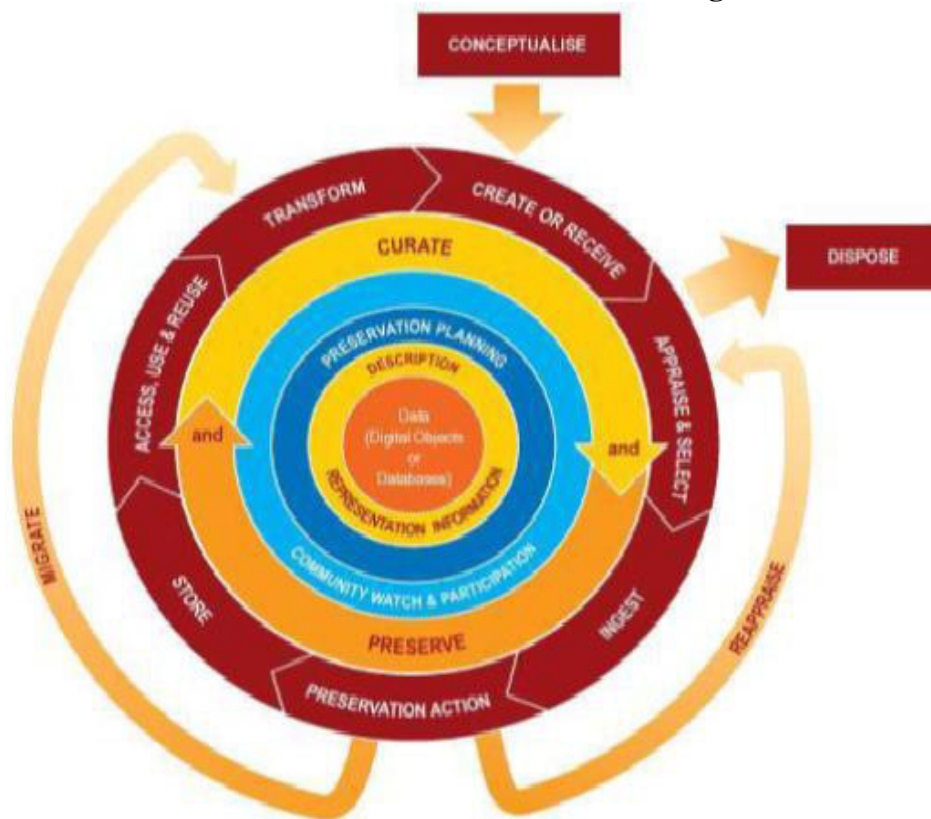
Fuente. Elaboración propia

Por lo descrito anteriormente se desprende que para el proceso de gestión de los datos es imprescindible que exista un plan de gestión de datos en los cuales el investigador es el que debe de estar totalmente comprometido para su selección y registro de acuerdo a sus necesidades, requerimientos y limitaciones.

La tesis de Valentín (2013) titulada *Gestión de datos de investigación* señala la definición de los datos de investigación por el National Science Board (NSBD) como: “Cualquier información que se puede almacenar en formato digital incluyendo texto, números, imágenes, vídeo, audio, software, algoritmos, ecuaciones, animaciones, modelos, simulaciones etc. Estos datos se podrán obtener por diversos medios, incluyendo la observación, el cálculo y la experimentación” (p.14). Explica que los datos para que puedan ser considerados como tales deben ser: datos numéricos (resultados de medidas de instrumentos, resultados de encuestas); resultados fruto de la medida de instrumentos; datos de encuestas; imágenes digitales (fotografías y diagramas); audios digitales (archivos sonoros, entrevistas); vídeos digitales; y documentación diversa (publicaciones, como notas de investigación de campo, bases de datos de modelos genéticos, descripciones.)

Valentín (2013) describe el Digital Curation Center (DCC), (ver Gráfico 2) como un “consorcio liderado por la universidad de Edimburgo y financiado por el Joint Information Systems Committee (JSIC)” (p.25), que nació como un centro para la preservación digital en el año 2013, sin embargo actualmente se encarga de apoyar a las instituciones del Reino Unido y del mundo a almacenar, gestionar y preservar los datos digitales, así como también se encarga del establecimiento de políticas en gestión de datos de investigación. Para esta institución los datos pueden ser objetos digitales (archivos de texto, archivos de imágenes, sonido etc. que se almacenan en formato digital) y bases de datos (colecciones estructuradas de registros o datos almacenados en un sistema informático). Este consorcio sugiere que en esta primera etapa es necesario que los responsables de la gestión de datos trabajen muy cerca con los investigadores ya que ellos son los que proporcionan la información necesaria de los datos con el fin de determinar el origen y el ciclo de vida de los mismos, así como también proporcionan los datos en una gran variedad de formatos. Por su parte, los responsables del proceso de gestión de los datos deben proporcionar un modelo o guía de gestión de los datos para ahorrar esfuerzos.

Gráfico 2 Ciclo de vida de los datos de investigación



Fuente. Digital Curation Center

Valentín (2013) continúa describiendo el modelo que presenta el DCC y que contiene una visión general de las etapas requeridas para realizar una gestión activa de los datos de investigación. Permite también definir roles y responsabilidades en la gestión de los datos de acuerdo con el ciclo de vida de los mismos. Los modelos que describe son: acciones aplicables a todo el ciclo de vida de los datos, ver tabla 1; acciones que pueden llevarse a cabo de forma secuencial ligadas al proceso de almacenamiento de los datos, ver tabla 2; y acciones ocasionales que se realizan cuando las circunstancias lo exigen, (ver tabla 3).

Tabla 1: Acciones aplicables a todo el ciclo de vida de los datos

Describir y representar la información	Asignar metadatos administrativos, descriptivos y estructurales utilizando las normas apropiadas para asegurar una descripción adecuada.
Preservar	Plan de mantenimiento para todo el ciclo de los datos. Incluye la creación de planes para la gestión de datos y para la correcta administración de todas las acciones del ciclo de vida.
Implementar programas de vigilancia y participación	Estar atento a las actividades de las comunidades de interés para la institución y realizar un seguimiento de las investigaciones realizadas.
Curar y preservar	Encargarse de la gestión y de la administración de todas las acciones previstas para promover la curación y preservación durante todo el ciclo de vida.

Fuente. Digital Curation Center

Tabla 2: Acciones secuenciales

Acciones Secuenciales	
Crear o recibir	Proceso de generación o recepción de los datos. Asegurarse de que los datos son recogidos en un formato adecuado y descritos con metadatos apropiados.
Evaluar y seleccionar	De acuerdo con las políticas establecidas se puede realizar una evaluación de los datos para seleccionar los que van a ser preservados a largo, mediano o corto plazo.
Desechar	Los datos pueden ser desechados o trasladados a otro archivo, depósito, centro de datos o destruidos de acuerdo a la política establecida.
Traspasar	Transferencia de datos para que sean reutilizados por otros investigadores.
Preservar	Acciones que se deben realizar para la preservación a largo plazo: acciones de corrección de errores, de limpieza de datos, validación de formatos, copias de seguridad etc.
Almacenar	Establecer un proceso de almacenamiento de datos sencillo y proporcionar apoyo y orientación siempre que sea posible. Automatizar los procesos y definir quién es el responsable de garantizar la calidad de los datos: el investigador, el gestor de información, etc.
Acceder, usar y reutilizar	Incluir metadatos, realizar indexaciones, controlar el acceso a los datos y establecer licencias y permisos.
Transformar	Crear nuevos datos a partir de los originales para crear un subconjunto de datos con el fin de crear nuevos resultados derivados. Mediante esta acción los datos vuelven a encontrarse al inicio de su ciclo de vida.

Fuente. Digital Curation Center

Tabla 3: Acciones ocasionales

Acciones ocasionales	
Reevaluar	Recuperar los datos debido a fallas en los procedimientos de validación para su posterior evaluación y re-selección.
Migrar	Migrar los datos a tecnologías o formatos más nuevos para: garantizar su preservación y evitar la obsolescencia del soporte físico o software; y conseguir una homogeneidad en el entorno de almacenamiento. En algunos casos es necesario preservar los datos originales para que el objeto digital siga siendo accesible y significativo.

Fuente. Digital Curation Center

Teniendo en cuenta todo lo descrito anteriormente se concluye que el plan de gestión de los datos es la parte más importante de todo el proceso. Se debe tener en cuenta el ciclo de vida de los datos, es decir desde su creación y almacenamiento hasta que se determine su obsolescencia y sean desechados pasando por un proceso de selección, conservando solamente aquellos que han servido como resultado para la investigación.

Según el Australian National Data Service (ANDS), citado por Valentín (2013) indica que "los datos de investigación deben pasar de una situación en la que son invisibles, inmanejables, desconectados y son de uso particular a convertirse en colecciones de datos estructurados, manejables, conectados y que pueden ser encontrados y reutilizados" (p.44). (Ver Gráfico 3). La clave para una gestión de datos de investigación exitosa es la planificación, organización y administración adecuada.

Gráfico 3 Transformación de los datos de investigación



Fuente. Australian National Data Service

Valentín (2013), en esta misma tesis, señala “que el ANDS desarrolla un programa de trabajo para la gestión de datos de la investigación, basado en 7 ejes principales:

1. **Formación y concienciación:** Es importante que los investigadores tengan conciencia de la importancia y las buenas prácticas de la gestión de datos de investigación, al permitir que estos estén disponibles para el intercambio, la validación y la reutilización. Para esta etapa la gestión de datos debe contar con las siguientes actividades: planificación, recopilación, análisis, publicación, archivo y reutilización.
2. **Elaboración de políticas para la gestión de datos de la investigación:** contar con políticas y procedimientos institucionales (directrices, protocolos y normas).
3. **Planificación de la gestión de los datos:** esta etapa es la más importante en este proceso. La responsabilidad principal de la gestión de los datos de

investigación recae en el investigador. La planificación incluye dentro de sus objetivos mejorar la toma de decisiones y por lo tanto se debe tener muy en cuenta los elementos que se deben considerar para planificar la gestión de los datos, tomando en consideración los tipos de datos que se crearán, que formatos se utilizarán, acceso a los datos, tipos de metadatos, roles y responsabilidades etc.

4. **Gestión de datos personales:** cuando la investigación se base en estos datos, los investigadores deben respetar las normas éticas establecidas a fin que puedan ser compartidos ética y legalmente.
5. **Licencias, copyright y datos:** hay que tener en cuenta los permisos para reutilizar los datos de acuerdo con las licencias existentes como por ejemplo Creative Commons.
6. **Captura de datos:** Desarrollo de infraestructuras o plataformas que permitan la integración de todos los procesos relacionados con la creación o captura de datos, la descripción, el almacenamiento y la compartición de los mismos.
7. **Almacenamiento de datos:** los materiales deben ser almacenados y conservados para poder justificar los resultados de la investigación y defenderlos si son desafiados, así como también permitir su reutilización especialmente cuando la investigación sea difícil o imposible de repetir.” Valentín (p.48, 49) (ver tabla 4).

Hernández-Pérez, y García-Moreno (2013), señalan en su artículo *Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios*, que “desde la aparición de internet y con el avance de las tecnologías de la información y la comunicación (TIC) se ha hablado mucho de diferentes siglas relacionadas con datos: big data, open data, linked data y sharing data , entre otras. Asimismo, existen diferentes tipos de datos: datos personales, sobre el contexto en el que vivimos, sobre nuestro ambiente, sobre objetos y productos, sobre procesos, etc. Existe también la creencia de que los datos por el simple hecho de estar en la web, pueden ser leídos o descargados, sin embargo, de acuerdo con la definición de la organización Open Definition Advisory Council, conformada por instituciones y representantes

destacados del movimiento de acceso abierto, los datos para ser considerados abiertos deben poseer algún tipo de licencia legal que permita utilizarlos, reutilizarlos y redistribuirlos” Hernández-Pérez (p. 260). Debido a que existen diferentes plataformas para procesar y almacenar datos, el autor indica la importancia de contar con una política y licencia legal para los datos abiertos al momento de implementar un repositorio de datos para el uso y reutilización de los mismos.

Estos mismos autores señalan que para la gestión de los datos de investigación se debe tener en cuenta los siguientes elementos:

1. **Planificación:** Es la parte más importante en la cual se marcarán los procesos y recursos para completar el ciclo de vida de los datos: objetivos del proyecto, gestión, explotación, políticas, plan de sostenibilidad de los datos y tipos de datos; cuándo, cómo y dónde se recogerán; formatos; metadatos obligatorios; políticas para compartir; cómo deberán ser citados; privacidad, seguridad y derechos de autor; costos etc., así como quién será el responsable de realizar cada una de estas actividades.
2. **Recolección:** Cómo será la recolección efectiva de datos de acuerdo a lo planificado: automática o manual. La recolección debería ajustarse a los formatos, vocabularios o, nombres que se hayan decidido.
3. **Control de calidad:** Mecanismos para prevenir errores al momento de registrar los datos. Puede consistir en realizar entradas dobles de datos, separadamente por dos personas, o en procedimientos y algoritmos que permiten detectar errores fácilmente.
4. **Descripción de datos:** Metadatos que permitan comprender el contenido, formato y contexto de datos (medidos, simulados, y procesados) que podrán ser diseminados para ser utilizados y reutilizados por los investigadores (catalogación con unos metadatos apropiados y acordados por la comunidad que los produce).
5. **Preservación:** En este proceso se asegurará la fiabilidad de los medios de almacenamiento, con políticas de verificación de la integridad, copias de seguridad y políticas de depuración a corto y largo plazo, donde los datos y los

resultados de sus análisis son almacenados para luego ser catalogados y diseminados.

6. **Descubrimiento:** Consta de dos partes: la primera hacer los datos realmente accesibles, fuera de ordenadores personales, en sitios web accesibles y en formatos reutilizables abiertos. La segunda, utilizar todas las herramientas posibles para que los datos puedan ser descubiertos (tesauros, ontologías), y, depositados en repositorios específicos.
7. **Integración o recuperación:** Uso de datos de múltiples fuentes, combinados de tal forma que se puedan analizar, es decir fuentes que traten del mismo tema, como por ejemplo para, construcción de viviendas, deberían poder combinarse con datos recolectados sobre la población o sobre equipamientos culturales. Una vez recuperados los datos, deben ser normalizados y recuperados para su análisis.
8. **Análisis:** Es la fase final y consiste en la generación de interpretaciones y visualizaciones para identificar patrones, así como también mostrar los resultados o descubrimientos de forma comprensible, es decir, procesar y analizar los datos en cualquier mecanismo, para que a partir de ellos se pueda construir información y posteriormente el conocimiento.” (p.262). (Ver Gráfico 4).

Gráfico 4 Plan de Gestión de datos de investigación



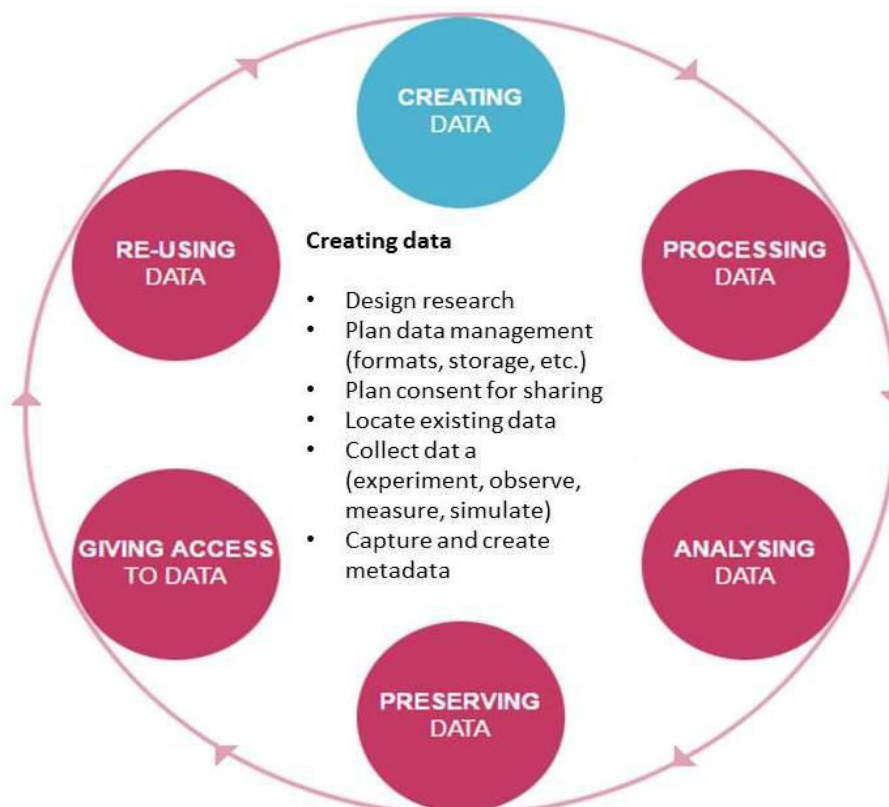
Fuente. Hernández-Pérez (2013).

En la tesis revisada de Pérez M. (2012) titulada *Una Aproximación al data curation y el rol del bibliotecario en su implantación: el caso de la Universitat Oberta de Catalunya*, señala “que los riesgos de la falta de gestión de datos son la creación de datos de forma indiscriminada y sin normalización ni control institucional, lo que conlleva un alto riesgo de pérdida de información a largo plazo, de escasa reutilización e interoperabilidad de los mismos”. Señala además (Pryor, 2012) “que para una correcta gestión de datos muchos organismos han propuesto una gestión de datos basándose en su ciclo de vida, lo cual permite asegurar que se desarrollen políticas y procesos, se definan roles y responsabilidades y se cree un marco tecnológico adecuado para su almacenamiento, es decir estándares y tecnologías a desarrollarse”.

Pérez, en esta investigación “señala que otro de los organismos que ofrece una propuesta de ciclo de vida de los documentos es el UD Data Archive, la cual define seis etapas para el ciclo de vida de los datos: (Ver gráfico 5).

1. **Creación de datos:** diseño de la investigación, creación de planes de gestión de datos, políticas de data sharing, localización de los datos, recolección y creación de metadatos.
2. **Procesamiento de datos:** introducción, digitalización, transcripción, traducción, validación, limpieza, descripción, almacenamiento y protección de datos (en los casos que deba asegurarse el anonimato)
3. **Análisis de los datos:** interpretación y preservación de los datos para su preservación.
4. **Preservación de los datos:** migración de los datos al formato recomendado, creación de backups, creación de metadatos y documentación, archivo de datos.
5. **Dar acceso a los datos:** distribuir, compartir y controlar el acceso a los datos, como establecer licencias y promocionar los datos.
6. **Reutilizar los datos:** realizar un seguimiento de la investigación, retomar nuevos resultados, enseñar y aprender de ellos”, (ver gráfico 5).

Gráfico 5 Etapas para el ciclo de vida de los datos según UD Data Archive



Fuente. <http://www.data-archive.ac.uk/create-manage/life-cycle>

Teniendo en consideración lo descrito anteriormente por los diferentes organismos al igual que la conclusión sobre la tesis de Valentín (2013), se concluye que el plan de gestión de los datos es la parte más importante en todo el proceso de gestión de los datos, así como la implementación de políticas de gestión de datos, teniendo en cuenta su ciclo de vida, es decir desde su creación y almacenamiento hasta determinar que son obsoletos, pasando por un proceso de selección de los datos, conservando solamente aquellos que han servido como resultado para la investigación.

Hemos revisado las dos propuestas sobre el Ciclo de Vida de los Datos de la DDC y de la UDK, mediante una tabla comparativa, con la finalidad de determinar las coincidencias y las diferencias entre ambas propuestas a fin de que nos sirva para determinar las características de la gestión de los datos de investigación de los investigadores. (Ver tabla 4).

Tabla 4: Cuadro comparativo del ciclo de vida de los datos: DDC y UD Data Archive

Digital Curation Center (DCC)		UK Data Archive	
Acciones secuenciales en el ciclo de vida de los datos		Etapas para el ciclo de vida de los datos	
Crear o recibir	Proceso de generación o recepción de los datos. Asegurarse de que los datos son recogidos en un formato adecuado y descritos con metadatos apropiados.	Creación de datos:	Diseño de la investigación, creación de planes de gestión de datos, políticas de data sharing, localización de los datos, recolección y creación de metadatos.
Preservar	Acciones que se deben realizar para la preservación a largo plazo: acciones de corrección de errores, de limpieza de datos, validación de formatos, copias de seguridad etc.	Preservación de los datos:	Migración de los datos al formato recomendado, creación de backups, creación de metadatos y documentación, archivo de datos.
Acceder, usar y reutilizar	Incluir metadatos, realizar indexaciones, controlar el acceso a los datos y establecer licencias y permisos.	Reutilizar los datos:	Realizar un seguimiento de la investigación, retomar nuevos resultados, enseñar y aprender de ellos.
Evaluar y seleccionar	De acuerdo a las políticas establecidas se realizar una evaluación de los datos para seleccionar los que van a ser preservados a largo, mediano o corto plazo.	Análisis de los datos:	Intepretación de los datos para su preservación
		Procesamiento de datos:	Introducción , digitalización, transcripción, traducción, validación, limpieza, descripción, almacenamiento y protección de datos (en los casos que deba asegurarse el anonimato)
Desechar	Los datos pueden ser desechados o trasladados a otro archivo, depósito, centro de datos o destruidos de acuerdo a la política establecida.		
Traspasar	Transferencia de datos para que sean reutilizados por otros investigadores.		
		Dar acceso a los datos:	Distribuir, compartir y controlar el acceso a los datos, como establecer licencias y promocionar los datos.
Almacenar	Establecer un proceso de almacenamiento de datos sencillo y proporcionar apoyo y orientación siempre que sea posible. Automatizar los procesos y definir quién es el responsable de garantizar la calidad de los datos: el investigador, el gestor de información, etc.		
Transformar	Creación de nuevos datos a partir de los originales a un formato diferente, para evitar la obsolescencia, creación de nuevos resultados derivados e iniciar el nuevo ciclo de vida de los datos.		

Fuente. Elaboración propia

Torres-Salinas, Robinson-García y Cabezas-Clavijo (2012), en su artículo *Compartir los datos de investigación en ciencia: introducción al data sharing*, explican que la iniciativa del acceso abierto a los datos no es algo nuevo. “De acuerdo a las investigaciones que se han realizado, se podría rastrear sus inicios en 1901 con Galton, quienes señalan que a Galton, a quién también se le asocia el nacimiento de la cienciometría (Godin 2007), se expresaba en la revista *Biometrika* en los siguientes términos -Pienso que nadie debiera publicar resultados biométricos sin depositar una copia de sus datos bien redactada y presentada en algún lugar donde todo aquel que lo deseara pudiera verificar su trabajo-.” Torres Salinas et al. (2012) señalan que “a partir de ese momento no solamente se habla de este tema, sino que se pone en marcha una forma de compartir y distribuir datos, con la creación de bancos de datos: Protein Data Bank (PDB) bajo los auspicios de la American Crystallographic Association, creando de esta manera una forma de compartir datos el cual ha mantenido un gran crecimiento desde sus inicios. Después de esta iniciativa han ido desarrollándose repositorios similares en diversas disciplinas: Neurodatabase (neurociencias); The cancer genome atlas (Oncología) o el Data observation network for Earth (ciencias de la tierra)”. (p.174). En todos estos repositorios existe el esfuerzo común (investigadores, empresas o sociedades de financiación y revistas científicas) para, no solamente acelerar el progreso de la ciencia sino desarrollar infraestructuras que permiten compartir un trabajo estándar de una disciplina.

En este mismo artículo, los autores señalan que “junto a la definición de las entidades del National Institutes of Health (NIH), existen intentos de hacer una taxonomía de los datos. La clasificación que ofrece la Research Information Network (RIN), establece tres criterios para determinar qué son los datos de investigación y cómo clasificarlos: (Ver Tabla 5).

Tabla 5 Clasificación de datos de investigación según la Research Information Network (RIN)

A. Definición de datos de los National Institutes of Health (NIH) <p>Por datos finales de investigación entendemos material factual registrado, aceptado por la comunidad científica y necesario para validar los resultados de la investigación.</p> <p>No son datos finales de investigación: notas de laboratorio, sets de datos parciales, análisis preliminares, borradores de trabajos científicos, planes para investigaciones futuras, informes que han tenido un proceso de revisión por pares, comunicaciones con colegas, u objetos físicos como genes o ejemplares de laboratorio.</p>	
B. Diferentes taxonomías de los datos de investigación	
B.1 Según el formato <ul style="list-style-type: none"> • Textos • Números • Imágenes • etc. 	B.2 Proceso de obtención Experimentales <ul style="list-style-type: none"> • Secuencias genéticas • Cromatografías Simulaciones <ul style="list-style-type: none"> • Modelos climáticos • Modelos económicos Observacionales <ul style="list-style-type: none"> • Encuestas • Experimentos
B.3. Según objetivo de recogida Específicos <ul style="list-style-type: none"> • Solo de interés para un proyecto de investigación. Alcance medio <ul style="list-style-type: none"> • De interés para una disciplina concreta De interés general <ul style="list-style-type: none"> • De interés para la ciencia en su conjunto e incluso de interés social 	B.4 Según fase de investigación Datos preliminares <ul style="list-style-type: none"> • Datos recién extraídos sin ningún tipo de procesamiento. • Denominados en inglés <i>rawdata</i>. Datos finales <ul style="list-style-type: none"> • Datos que ya han sido procesados y combinados con otros. Denominados en inglés <i>final research data</i>.
C. Medios para compartir los datos de investigación	
C1. Comunicación formal Repositorios de datos/bancos de datos <ul style="list-style-type: none"> • Centralizados • Descentralizados • Federados • Ciberestructuras 	C.2 Comunicación informal A petición <ul style="list-style-type: none"> • Por ejemplo vía correo electrónico Descentralizada <ul style="list-style-type: none"> • Por ejemplo a través de webs personales

Fuente. Torres-Salinas, et.al. (2012)

2.3.2. Preservación de datos de investigación

Millán-González, Saorín, Ferrer-Sapena, Aleixander-Benavent y Peset (2013) en su artículo *Gestión de datos de investigación: infraestructura para su difusión*, señalan que “preservación digital es el proceso de establecer y mantener a largo plazo repositorios digitales de referencia capaces de salvaguardar los datos actuales y futuros. Otros la definen como la administración de datos científicos digitales a los que

se proporciona un valor añadido dándoles un contexto y una vinculación a su origen de manera que se facilite su reutilización, integración y promoción. Seguidamente señalan que para la preservación y la reutilización de datos no solo hay que resolver aspectos técnicos, sino también de organización y de procedimiento económico, financieros y de personal, administración de su propiedad, obligaciones legales, requisitos de auditoría, restricciones sociales de uso etc. Este conjunto de acciones configura lo que se denomina *data curation*, actividad que recibió un gran impulso con la creación en marzo de 2004 del Digital Curation Centre (DCC) en Gran Bretaña” (p. 417). De acuerdo con lo señalado por los autores, es necesario que al momento de realizar un plan de gestión de datos, se tome en cuenta como uno de los aspectos más importantes la preservación de los datos. Para lograrlo es necesario que los gestores de la información, debido al perfil que los caracteriza, cambien sus roles hacia *data curators*, realizando actividades que implican la selección, registro, clasificación y almacenamiento de los datos con el fin que estos sean reutilizados. Cabe indicar que todas estas actividades deben estar totalmente coordinadas con los propios investigadores.

Pérez (2012) en su tesis titulada *Aproximación al data curation y el rol del bibliotecario en su implantación: el caso de la Universitat Oberta de Catalunya* “amplía un poco más la definición del Digital Curation Centre sobre *data curation*, señalando que Ross Harvel (2010), describe que, “el Data Curation se encarga de la gestión activa de los datos durante el tiempo que siguen teniendo interés académico, científico, administrativo y personal, con el objetivo de favorecer su reproducción, su reutilización y agregándoles valor, los datos se gestionan desde su creación hasta que se determina que ya no son útiles, su accesibilidad a largo plazo, su conservación, su autenticidad y su integridad”. Así mismo, señala que en el término anglosajón *data curation* se utiliza para referirse a las acciones necesarias para el mantenimiento y utilización de los datos digitales y los resultados de investigación durante todo su ciclo de vida y desde su generación”. (p.17). En consecuencia podemos decir que *data curation* representa a un conjunto de acciones para el mantenimiento y utilización de los datos digitales y los resultados de investigación. En realidad nos referimos a una función dentro del rol del gestor de la información que tiene como fin agregar valor a

los datos para mantener el interés académico, científico y administrativo, antes de la creación de los datos, durante su creación y con el uso a lo largo de su ciclo de vida.

Valentín (2013), señala que “actualmente se ha producido un cambio en los criterios tradicionales para la preservación a largo plazo, según el Digital Preservation Coalition (DPC), actualmente se tienen en cuenta los siguientes criterios:

- Que el formato del recurso sea legible actualmente y en un futuro:
- Que el recurso esté en un soporte gestionable para su transferencia y/o almacenamiento
- Que la institución tenga pleno derecho a manipular los datos para asegurar su acceso en entornos informáticos del futuro, no se puede preservar un recurso reproduciéndolo o reformatéandolo si no se tiene el permiso del autor.
- Que el recurso disponga de documentación, incluyendo los metadatos”. (p.23).

Podemos decir que se deben seguir los criterios de cambio sobre la preservación proyectados al futuro, considerando formatos, características de recurso y soportes. Otro aspecto a tomar en cuenta es que las organizaciones deben seguir los criterios para generar entornos informáticos del futuro incluyendo los permisos de los investigadores relacionados a sus datos de investigación.

2.3.3. Políticas y lineamientos para la gestión de los datos de investigación

2.3.3.1. Políticas y lineamientos para la gestión de datos de investigación a nivel internacional

Gómez, Méndez y Hernández-Pérez, (2016) *Datos y metadatos de investigación en ciencias sociales y humanidades: una aproximación desde los repositorios temáticos de datos*, señalan que hay lineamientos recomendados por organizaciones gubernamentales de diferentes países que están considerando la importancia de compartir datos de investigación como “ la OCDE (2015), lo exige el Gobierno de Estados Unidos desde 2013 a través de las diversas agencias de financiación: National Science Foundation (NSF,2014), y los National Institutes of Health (NIH, 2015), entre

otros. En Europa, el acceso abierto a los datos de investigación ha sido, hasta ahora, sólo un piloto (ORD Pilot) para nueve áreas de proyectos financiados en el marco de Horizon 2020 invitando a otras áreas y programas a participar voluntariamente (European Commission, 2016). Sin embargo, el 19 de abril de 2016, la Comisión declaró que los datos de investigación abiertos serán la opción por defecto para todos los nuevos proyectos financiados en H2020 a partir de 2017 (COM 2016, p.8)” (p.546). En este Programa se señala que “en este proyecto piloto se debe proporcionar acceso abierto a los datos necesarios para validar los resultados presentados en las publicaciones científicas. Los requisitos legales para los proyectos que participan en este piloto están establecidos en el Artículo 29.3 del Convenio de Donación: el beneficiario debe depositar los datos en un repositorio de datos de investigación para que sea posible acceder, explotar, reproducir y difundir los datos de forma gratuita para cualquier usuario. Así mismo, manifiesta que una vez iniciado el proyecto los beneficiarios de los mismos deben formular un Plan de Gestión de Datos (DMP) para definir qué conjuntos de datos generará o procesará, cómo estos datos serán accesibles, curados, almacenados y conservados.

De acuerdo con lo descrito en las Directrices de la European Research Council (2016), en su versión en inglés, *Guidelines on the implementation of Open Access to Scientific Publications and Research Data in Projects supported by the European Research Council under Horizon 2020*, se señala que “el Consejo Europeo de Investigación (ERC) apoya el principio básico de acceso abierto a los datos de investigación. Por lo tanto, recomienda a todos sus investigadores financiados que sigan las mejores prácticas al retener los archivos de todos los datos de investigación que han producido y utilizado durante su trabajo y que estén dispuestos a compartir estos datos con otros investigadores siempre que no estén vinculados por restricciones de derechos de autor, requisitos de confidencialidad o cláusulas contractuales”.

Por lo expuesto en las Directrices del ERC dentro del marco del Programa Horizonte 2020 podemos deducir que en la Unión Europea ya existen lineamientos específicos para la presentación de proyectos sobre datos abiertos de investigación según normas establecidas para la investigación e innovación a los cuales deben ceñirse todos los investigadores. En estas directrices se señalan además, tres pasos a

seguir sobre el mandato de acceso abierto a publicaciones científicas: depósito de publicaciones en repositorios; selección de la ruta de acceso abierto (acceso verde o dorado); y acceso abierto a las publicaciones. En el primer paso: depósito de publicaciones en repositorios se especifica que el investigador beneficiario del proyecto debe tratar de depositar al mismo tiempo que la publicación, los datos de investigación necesarios para validar los resultados presentados en las publicaciones científicas depositadas, idealmente en un repositorio de datos. Así mismo, recomienda utilizar repositorios específicos para sus publicaciones como por ejemplo (los mencionados en la tabla 6).

Tabla 6: Repositorios específicos para publicaciones

Mandato de acceso abierto a publicaciones científicas	
Tema	Repositorio
1. Ciencias de la vida	Europa PubMed
2. Ciencias físicas e ingeniería	arXiv
3. Monografías, capítulos de libros y otros textos largos	OAPEN
4. Recursos específicos de genoma humano. Integra información de una amplia gama de especies.	GENE
5. Datos estadísticos de cómo se administra la atención médica en los Estados Unidos.	Agency for Healthcare Research and Quality, Data & Surveys (AHRQ)
6. Si no hay un repositorio apropiado de disciplina específica-	Repositorios institucionales o en centros centralizados como Zenodo

Fuente. Elaboración propia.

Cabe indicar que no recomiendan como repositorios: Research Gate, Academia.edu; páginas web personales, institucionales o el depósito en una cuenta de Dropbox. De acuerdo con lo descrito podemos concluir que es recomendable que los datos de investigación deben ser depositados al mismo tiempo de la publicación en un repositorio recomendado según a la disciplina que se investigue, en repositorios institucionales o en centros centralizados para validar los resultados presentados en las publicaciones científicas. Así mismo deben almacenarse en formatos de archivos estandarizados a fin que puedan ser reutilizados.

Arquero, y Marco (2014) en *El Portal de datos abiertos de la Unión Europea: análisis y evaluación*, (Ver Gráfico 6) señalan que “las iniciativas de datos abiertos constituyen un elemento clave que favorece las políticas de transparencia de las Administraciones Públicas, aseguran el libre acceso a los datos a todos los ciudadanos y promueven oportunidades de negocio a partir de la reutilización de los datos públicos como materia prima. Estas iniciativas se han apoyado en los últimos años en dos pilares clave: *el desarrollo de las nuevas tecnologías* que facilitan que la información pueda ser publicada, relacionada y compartida de forma eficaz; y *las políticas de gobierno abierto u open government*, que fomentan estrategias orientadas al avance de una sociedad digital asentada sobre los principios de transparencia, colaboración y participación” (p.100).

En este mismo artículo los autores señalan que el desarrollo de los portales de datos abiertos es el elemento más visible de las políticas en materias de Gobierno Abierto y una de las principales herramientas de acceso a los recursos informativos. Estos portales permiten que el conjunto de datos que se pone a disposición de los ciudadanos sean mantenidos, se encuentren actualizados y utilicen formatos estructurados que permitan su consulta y reutilización desde un solo lugar evitando la dispersión de los contenidos entre las diferentes páginas web de los organismos públicos” (p.101). Por lo expuesto, se identifica que es necesario desarrollar portales de datos abiertos en una sola plataforma a fin de facilitar el acceso a conjuntos de datos que pueden ser reutilizados para fomentar nuevas investigaciones e innovaciones y facilitar la transparencia y la participación social de los ciudadanos.

Valentín (2013) señala que “tomando como base el Australian Code for the Responsible Conduct of Research, quien es el encargado de orientar a las instituciones y a los investigadores australianos para que realicen prácticas de investigación,” (p.41) en su segundo capítulo “presenta políticas referentes a la gestión de los datos de investigación, su almacenamiento, su mantenimiento más allá del final del proyecto, y todo lo relativo al acceso a ellos por parte de la comunidad científica” (p.41). Este código recomienda que el periodo durante el cual deben conservarse los datos debe ser determinado por el tipo específico de la investigación y por el propio investigador. Por su parte, la institución debe garantizar la seguridad y confidencialidad de los datos y

por lo tanto debe tener bien definida una política sobre la propiedad y el acceso a los mismos.

Gráfico 6: Portal de Datos Abiertos de la Unión Europea



Fuente. Arquero, A. (2014)

El autor señala además, “que el Australian Nacional Data Service (ANDS), ente coordinador australiano sobre la gestión de datos de investigación, considera que toda política debe contar con lineamientos sobre como registrar los datos” (p.50). Por lo expuesto, podemos mencionar que es necesario establecer una política de gestión de datos de investigación al momento de elaborar el plan de gestión de datos, a fin que los investigadores estén informados; así mismo, se podría tomar como modelo otras consideraciones sugeridas en el plan de políticas de la ANDS. Cabe indicar, que es recomendable que el establecimiento de los lineamientos de la política se debe desarrollar junto con los investigadores, dado que estos son los responsables de proporcionar los datos de investigación.

Ferrer-Sapena, Peset y Aleixandre-Benavent (2011), señalan que otra iniciativa de open data “se enmarca en el contexto de gobierno abierto y su filosofía, que es la del acceso abierto a determinados datos sin restricciones de copyright” (p.262). Describen diversos lineamientos de open data en el mundo, como por ejemplo: en Estados Unidos, se dio la primera medida formulada por Barak Obama en enero de 2009 en el Memorándum de Transparencia y Gobierno Abierto, quien señaló que: “un solo bit de información puede generar un torrente de creatividad” (p.264). En esta política se señala que el acceso a la información pública es un derecho, y cualquier documento creado por el gobierno debe ser de dominio público sin restricciones de derechos de autor y, en este sentido, el gobierno americano dispone de más de 380,000 data sets (p.264). Al poco tiempo países como Reino Unido, Nueva Zelanda y Canadá siguieron una tendencia de apertura de datos a través de sus propios portales. En el Reino Unido se dispone de más de 6.900 datasets que ofrecen información sobre la política del gobierno y, permiten que los ciudadanos entiendan cómo funcionan y cómo se formulan sus políticas. Un ejemplo, de estos *datasets* se muestran en el sitio web “*Where does my money go?*”, que muestra dónde va el dinero público (p.265). Otras normativas o leyes sobre open data se encuentran por ejemplo en España en la Ley 37/2007 sobre *Reutilización de la Información del sector Público*, que regula y fomenta la utilización de los datos generados y custodiados por la administración pública. Otro ejemplo en España es el *Plan Avanza* (2009-2012) que ha fomentado este acceso mediante el *Proyecto Aporta*, que tiene como objetivo principal incrementar la cobertura y conectividad de la Red en España y, así mismo, “pretende conseguir productos y servicios TIC avanzados”: proyecto e-Administración” (p.266). A partir de todas estas iniciativas y lineamientos a nivel mundial sobre acceso abierto, podemos mencionar que en América Latina y en el Perú se han empezado a implementar lineamientos sobre gobierno abierto. En el Perú, en el año 2013 se promulgó la Ley 30035, que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación del Acceso Abierto.

En el Seminario internacional “*desarrollando innovación en transparencia: datos abiertos en el marco de los estándares de la OCDE*” (Perú, Presidencia del Consejo de Ministros (2016), se señala que países como Reino Unido y Estados Unidos se han enfocado en el desarrollo de los datos abiertos y encabezan el Open Data

Barometer 2014. Esta es una iniciativa de una clasificación mundial impulsada por la World Wide Web Foundation y el Open Data Institute del Reino Unido. Se encarga de calificar a los países en base al nivel de publicación de datos gubernamentales respecto a su beneficio e impacto. Este barómetro incluye también a los países latinoamericanos, estando Chile en el puesto 15, Brasil en el 21, México en el 24, Uruguay en el 25, y Perú en el puesto 33.

2.3.3.2. Políticas y lineamientos para la gestión de datos de investigación en América Latina.

En los últimos años se ha empezado a implementar en América Latina políticas y lineamientos sobre Gobierno Abierto y, en consecuencia, sobre gestión de datos abiertos. En el año 2011, se lanzó la iniciativa Alianza para el Gobierno Abierto – AGA- (Open Government Partnership - OGP) con el fin de apoyar el desarrollo de gobierno abierto en los países. AGA es una iniciativa global que tiene como objetivos mejorar los niveles de transparencia mediante la apertura de datos que permita a la administración pública a rendir cuentas y a la ciudadanía ejercer un adecuado control social, así como permitir su participación en el diseño e implementación de las políticas públicas, favoreciendo de esta manera espacios de colaboración entre las entidades públicas y la sociedad civil. Actualmente son 69 países los que conforman esta Alianza. Para lograr estos objetivos, OGP reúne a gobiernos y organizaciones de la sociedad civil para que trabajen como socios. La presidencia de OGP cuenta con representantes tanto de gobiernos como de sociedad civil.

La Organización para la Cooperación y el Desarrollo Económico (OCDE), en su publicación *Gobierno Abierto en América Latina. Estudios de la OCDE* (2015) realizó un estudio entre los años 2013 y 2014 sobre “estrategias de gobierno abierto, acceso a la información y datos abiertos en América Latina.

En este mismo estudio la OCDE (2015), en la sección *Evaluación y Recomendaciones* se considera que “los países de América Latina y el Caribe, deberían formular políticas de datos abiertos de la mano con las acciones significativas emprendidas hasta ahora para asegurar el acceso a la información. Esto reforzaría su

reconocimiento y no solo ayudaría a aumentar los beneficios de los datos abiertos para los elementos clave de gobernanza tales como la transparencia y la apertura sino también serviría como motor de generación del valor social y económico para el sector público, la economía en su conjunto y la sociedad en general” (p.26). Según este mismo estudio, resulta imperativo pasar de un enfoque de oferta de datos a una perspectiva de demanda de datos que favorezca la reutilización de estos y la creación de valor, lo que supone ampliar el suministro de datos en base a las necesidades de los usuarios: datos geográficos, medioambientales y educativos.

Continúa señalando que para alcanzar los objetivos de datos abiertos se requieren actividades de capacitación que busquen separar las brechas de habilidades o competencias. En este sentido los países de América Latina, deberían considerar reforzar la cooperación con el sector académico a fin de que provean la capacitación en datos abiertos y desarrolle herramientas (seminarios, manuales sobre datos abiertos de gobierno, directrices sobre estándares de calidad de datos). Así mismo considera que deberían desarrollar portales centralizados de datos abierto a través de una ventanilla única”. (OCDE, 2015, p.26). De acuerdo con lo expuesto, señalamos que es necesario que se desarrolle un portal único y centralizado en América Latina sobre datos abiertos, a fin de fortalecer la transparencia y la participación social de los ciudadanos. Así mismo, se debería tomar en cuenta la recomendación de la OCDE en este estudio sobre la conveniencia de reforzar la cooperación entre el Gobierno y las instituciones académicas para lograr los fines deseados.

En el *Seminario internacional “Desarrollando innovación en transparencia: datos abiertos en el marco de los estándares de la OCDE”*, Perú. Presidencia del Consejo de Ministros (2016) se señala que “en la mayoría de los países de América Latina, son significativos los avances en políticas de acceso a la información y la creación de leyes para lograrlo. Algunos han creado webs y aplicaciones para la transparencia y la lucha contra la corrupción, y otros han instaurado políticas de participación ciudadana. Sin embargo, pocos han acogido del todo la iniciativa de datos abiertos” (p.17). Hace hincapié en que los datos abiertos implican más que simplemente revelar información pues apuntan a que diversos actores, como la

sociedad civil y el sector privado, tengan acceso a data que les permita generar nuevos productos, información o aplicaciones.

Otra iniciativa sobre acceso abierto en Latinoamérica es “LA Referencia: Red Federada de Repositorios Institucionales de Publicaciones Científicas, presentada por la Red Clara. En sus inicios (2010-2012) fue financiada por el Banco Interamericano de Desarrollo (BID). En su página web la Red Clara señala que LA Referencia “es una iniciativa regional que busca compartir y dar visibilidad a la producción científica generada en las instituciones de educación superior y de investigación científica en América Latina, gracias a la creación de un marco de acuerdos técnicos y organizativos para construir la Red Federada de Repositorios Institucionales con el fin de generar mecanismos planificados para el Acceso Abierto al conocimiento científico.”

Nakano y Azlirevich (2016), señala que “son los gobiernos que forman parte de LA Referencia los que deciden que esta red funcione y determinan cuáles son sus objetivos e iniciativas. Manifiestan que la iniciativa regional sobre datos primarios se empezó a trabajar este año por grupo de trabajo de repositorio de datos. En esta reunión se trabajó en el consenso sobre el alcance de la función de La Referencia, los objetivos y definiciones, así como la identificación de los principales iniciativas de repositorios de datos de la región, basándose en el diagnóstico de una encuesta que se hizo el año pasado, además de aspectos identificados en distintos repositorios como son Open Doar, Coar, y Red3data. Así mismo identificaron los esquemas de metadatos para datos primarios de investigación de mayor relevancia; el análisis de la literatura y esquemas de metadatos posibles; además se realizaron entrevistas con actores como: DataCite; Scielo; Redalyc, OpenAire, entre otros. En el análisis de identificó que si bien existen lineamientos en la región en Argentina, México, Perú y Chile, los repositorios de datos encontrados, algunas veces no son repositorios de datos como tales, sino portales o sitios web, mientras que la mayoría trabaja en la visualización de los datos que cubren datos de gestión y gobierno, pero no datos científicos.

Señala, además, que en el 2017 se empezará a implementar un piloto de repositorio de datos que tenga como requisitos mínimos: el protocolo OAI-PMH; y

esquemas de metadatos para facilitar la interoperabilidad; además, se buscará un esquema común para publicaciones y datos.

2.3.3.3. Políticas y lineamientos para la gestión de los datos de investigación en el Perú.

En el *Seminario internacional “Desarrollando innovación en transparencia: datos abiertos en el marco de los estándares OCDE”* (Perú. Presidencia del Consejo de Ministros (2016) se señala que la OCDE considera los datos abiertos como un mecanismo que trasciende a la difusión de la información y que el Perú, como aspirante a ser miembro de este organismo, está poniendo énfasis en adecuar sus estándares a los exigidos por la organización, como por ejemplo establecer una política de datos abiertos y gobierno abierto. En relación con esta iniciativa, la OCDE señala que “el Perú se está esforzando por impulsar los datos abiertos y la agenda de transparencia del gobierno; y resalta que en el diseño del Plan de Gobierno Abierto se menciona a la información pública como un activo clave para fortalecer la transparencia y la participación. Resalta también que el Estado considera a los datos abiertos como un factor crítico para aumentar el acceso público a sus acciones y decisiones, para mejorar la transparencia (p.20). Así mismo, señala que los datos abiertos siguen teniendo una modesta presencia en las agendas del gobierno, del sector privado y de la sociedad civil y recomienda impulsar las políticas de datos abiertos en los tres espacios, pues la sinergia ayudará a que los datos abiertos tengan un real valor económico y social” (p.20).

Gráfico 7 Significado de Datos Abiertos



Fuente. Perú. Presidencia del Consejo de Ministros (2016). Seminario Internacional. PCM. Perú.

En abril del año 2012, el Perú se incorporó a la Open Government Partnership (OGP), iniciativa denominada Alianza para el Gobierno Abierto (AGA), al ser aprobado un Plan de Acción que contenía compromisos concretos en materia de transparencia y acceso a la información pública, integridad pública, gobierno electrónico y promoción del desarrollo de nuevas tecnologías, participación ciudadana y rendición de cuentas. Dentro de estos compromisos se encuentra la presentación de la evaluación anual a dicha organización y la elaboración de dos Planes de Acción de Gobierno Abierto, el de 2012-2014 y el de 2015- 2016 (actualmente en ejecución), a cargo de la coordinación de la Secretaría de Gestión Pública (SGP), órgano de línea de la Presidencia del Consejo de Ministros y rector del Sistema Nacional de Modernización de la Gestión Pública. Luego de esta incorporación se creó la Comisión Multisectorial para el seguimiento de la implementación del Plan de Acción del Gobierno Abierto. En su Plan de Acción de Gobierno Abierto 2015-2016, el Perú reafirma el compromiso para cumplir con los objetivos de la OGP y precisa que el Gobierno Abierto es un eje transversal de la Política Nacional de Modernización de la Gestión Pública, y aprueba mediante la promulgación del Decreto Supremo N° 004-2013-PCM, publicado el 6 de enero del 2013. Esta comisión está adscrita a la Presidencia del Consejo de Ministros y está conformada por el Estado, la sociedad civil y los gremios empresariales. Con fecha 12 de febrero de 2017, se promulga el Decreto Supremo N° 016-217-PCM, que aprueba la “Estrategia nacional de datos abiertos gubernamentales del Perú 2017-2021” y el “Modelo de datos abiertos

gubernamentales del Perú”. Este decreto tiene como finalidad promover la apertura de datos abiertos de las entidades públicas, la innovación en la generación del valor público con la reutilización de los datos abiertos para la creación de nuevos productos y servicios con el uso de las tecnologías de la información y de comunicación (TIC), para el desarrollo social y económico del país, en el marco de un gobierno abierto.

Arobes, en el *Seminario internacional “Desarrollando innovación en transparencia: datos abiertos en el marco de los estándares OCDE”* (Perú. Presidencia del Consejo de Ministros (2016), señala que la Secretaría de Gestión Pública (SGP) de la Presidencia del Consejo de Ministros, resalta que debido a que el trabajo en datos abiertos en nuestro país, se ha desarrollado a nivel de gobiernos locales (municipios) y luego ha escalado al gobierno nacional; la estrategia nacional adoptada para los datos abiertos se ha construido en base a experiencias exitosas internacionales y locales. La estrategia nacional comprende tres líneas de acción: establecimiento del marco normativo e institucional; el desarrollo de infraestructura y la promoción de iniciativas y espacios de participación. Para el proceso de la apertura de datos a nivel nacional la SGP ha elaborado una guía rápida de apertura que va a permitir identificar los datos a liberar, hacer un inventario de ellos, priorizarlos y ofrecer un catálogo de datos. Así mismo, se ha diseñado el Portal Nacional de Datos Abiertos (ver gráfico 8) (www.datosabiertos.gob.pe) de acuerdo a buenas prácticas internacionales bajo los estándares del Plan Nacional de Gobierno Electrónico. Dentro de este portal se permitirá el libre acceso y descarga de datos pero tendrá permisos de modificación restringidos para que no se manipule deliberadamente la información. En relación con el perfil de los metadatos cumple con la norma ISO 19139 Geographic Information y la norma ISO 15836 Dublin Core Metadatos.

Gráfico 8 Portal Nacional de Datos Abiertos



Fuente. Perú. Presidencia del Consejo de Ministros (2016).

Finalmente, la OCDE señala que para lograr el éxito en un plan de datos abiertos es necesario que los gobiernos cumplan cuatro funciones específicas: deben ser proveedores de datos; asumir el liderazgo de la política de datos abiertos; ser catalizadores del uso de los datos; y también deben ser usuarios.

Con fecha 15 de mayo de 2013, se promulgó la Ley N°30035, que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, publicada en el Diario El Peruano el 5 de junio de 2013. El 24 de enero de 2015 se aprobó su Reglamento Abierto” que señala en el artículo 3.2 “las entidades privadas o personas naturales, cuya producción intelectual haya sido financiada o realizada total o parcialmente utilizando fondos y/o subvenciones del Estado. Sin perjuicio de lo establecido en los numerales precedentes, también podrán incorporarse voluntariamente al Repositorio Nacional Digital, las obras, los datos procesados y las estadísticas de monitoreo... “Así mismo, en el artículo 4.1 se “establece que el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto es administrado por el Consejo Nacional de Ciencias, Tecnología e Innovación Tecnológica (CONCYTEC), y se constituye en la máxima instancia para recolectar,

integrar, estandarizar, almacenar, preservar y difundir la producción nacional de ciencias, tecnología y innovación en los repositorios”. (Ver gráfico 9).

Gráfico 9 Portal del Repositorio Digital ALICIA



Fuente. CONCYTEC (2016).

En junio de 2016 el Consejo Nacional de Ciencias Tecnología e Innovación Tecnológica (CONCYTEC) promulgó la Directiva N° 0004-2016- CONCYTEC-DEGC, la cual regula el procesamiento de información en los repositorios institucionales. Allí se señala que el Repositorio Nacional Digital de Ciencias, Tecnología e Innovación de Acceso Abierto – ALICIA (www.alicia.concytec.gob.pe) se define como el sitio centralizado en el que se mantiene información digital que es resultado de la producción en ciencia, tecnología e innovación (libros, publicaciones, artículos de revistas especializadas, trabajos técnico-científicos, programas informáticos, datos procesados y estadísticas de monitoreo, tesis académicas y similares), y contribuye a conservar, preservar y ofrecer acceso abierto a la producción científica nacional, poniéndola a disposición de la comunidad académica, científica y la sociedad en general. En estas Directrices se señala que el objetivo de ALICIA es conformar una red interoperable de repositorios institucionales, a partir del establecimiento de políticas, estándares y protocolos para el intercambio de información comunes a todos los integrantes de la red.

Dentro de los lineamientos establecidos en las Directivas del CONCYTEC, señalados anteriormente, se considera al Directorio Open Doar como un espacio donde deben estar registrados los repositorios institucionales en el Perú. Open Doar es una iniciativa sobre directorios en acceso abierto. Esta iniciativa fue creada por la Universidad de Nothingam en Gran Bretaña. La característica de este Directorio es que todos los contenidos son validados antes de ser publicados con el fin de validar la información registrada en el repositorio, lo cual constituye un punto clave para que la mayoría de instituciones consideren elegir a este directorio para publicar sus repositorios. El Directorio Open Doar presenta una lista de repositorios a nivel mundial (ver gráfico 10) y permite realizar búsquedas de contenidos por temas, tipo de contenidos ((ver gráfico 11), tipos de softwares (ver gráfico 12), países, idioma, lo cual resulta muy útil para buscar repositorios especializados.

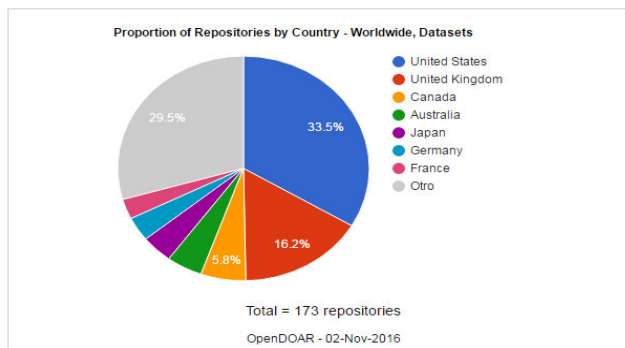
Gráfico 10 Proporción de repositorios de datos por países según Open Doar

OpenDOAR

Directory of Open Access Repositories

[Home](#) | [Find](#) | [Suggest](#) | [Tools](#) | [FAQ](#) | [About](#) | [Contact Us](#)

Proportion of Repositories by Country - Worldwide, Datasets



This chart is based on the number of repositories in each Country. However, some organisations have two or more repositories - over 20 in some cases - and this arguably skews the results.

For a different viewpoint, please see the equivalent chart for [Repository Organisations](#), in which each organisation only counts once, regardless of how many repositories it hosts.

For further data, please see the corresponding [table of repositories](#) sorted by country.

[Show embedding code](#)

[Show legacy chart and embedding code](#)

© 2016, [University of Nottingham, UK](#). Last updated: 02-Nov-2016

Fuente. Open Doar

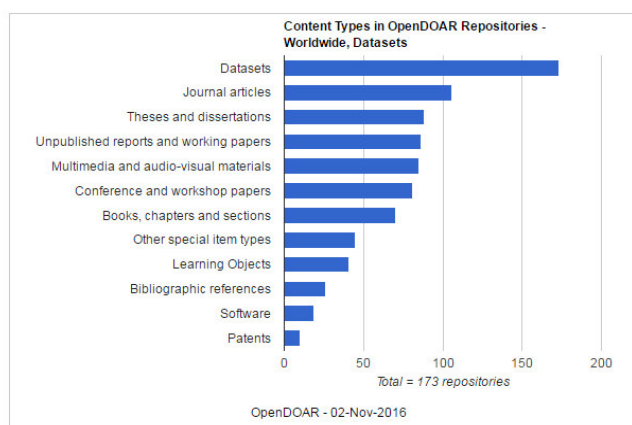
Gráfico 11 Tipos de contenidos de datos en repositorios según Open Doar

OpenDOAR

Directory of Open Access Repositories

[Home](#) | [Find](#) | [Suggest](#) | [Tools](#) | [FAQ](#) | [About](#) | [Contact Us](#)

Content Types in OpenDOAR Repositories - Worldwide, Datasets



N.b. Most repositories hold several Content Types.

[Show embedding code](#)

[Show legacy chart and embedding code](#)

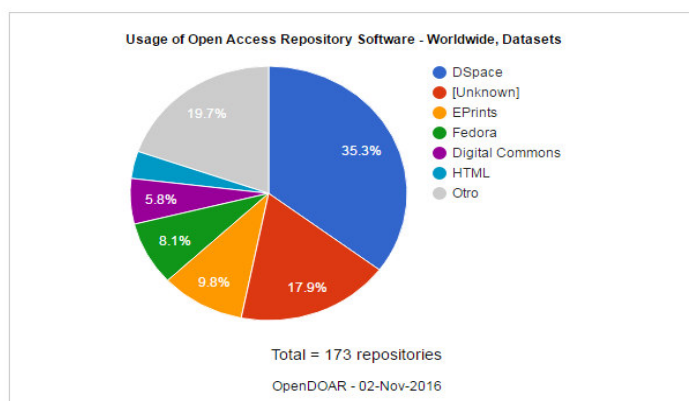
Fuente. Open Doar

Gráfico 12 Tipos de softwares utilizados en repositorio de datos según Open Doar

OpenDOAR

Directory of Open Access Repositories
Home | Find | Suggest | Tools | FAQ | About | Contact Us

Usage of Open Access Repository Software - Worldwide, Datasets



For further data, please see the corresponding [table of repositories](#) sorted by software platform.

[Show embedding code](#)

[Show legacy chart and embedding code](#)

Fuente. Open Doar

El CONCYTEC presenta un portal de datos abiertos (Perú. Presidencia del Consejo de Ministros 2016), (Ver gráfico 13), en el cual se encuentra una definición y lineamientos sobre datos abiertos, así mismo se presenta un directorio de instituciones que tienen portales de datos abiertos. Este portal presenta las siguientes características: se muestra un enlace a las páginas web de algunas instituciones a nivel del gobierno central tales como: Congreso de la República, Academia de la Magistratura, las Cortes Supremas de Justicia, entre otras instituciones; se presenta enlaces a universidades nacionales, sin embargo, en relación a los datos presentados, solamente están relacionados a las funciones que se realiza en cada una de las instituciones y con características propias de cada institución. En relación a la investigación en universidades, el portal presenta conjuntos de *datasets* generados por el Directorio de Investigadores e Innovadores tales como: cantidad de investigadores que estudiaron en universidades públicas por grados académicos; cantidad de investigadores que trabajan en universidad públicas por área temática; (OCDE) y cantidad de investigadores que estudian por áreas temáticas (OCDE). Cabe indicar que estos datos

se encuentran publicados al 10 de julio de 2015. Sin embargo, no se presentan portales de datos de investigación de cada una de las universidades señaladas.

Gráfico 13 Concytec. Portal de datos abiertos



Fuente. CONCYTEC (2016).

Mori, (2013) en el informe *Hacia una universidad transparente: informe sobre la transparencia y el acceso a la información en las universidades públicas del Perú*, describe en el capítulo sobre la investigación científica y canon minero “que en el 2011. las universidades públicas recibieron por canon minero S/. 926 millones para usarlos en investigación científica y tecnológica, pero solo utilizaron el 9% del canon. Señala que esto se puede deber, a que la transferencia del canon para las universidades se decidió sin conocer sus capacidades reales para invertirlos; además, señala que hay que tomar en cuenta que el país no tiene un sistema nacional de investigación que esté articulado con los proyectos universitarios. Señala Mori (2013), que de acuerdo al Ministerio de Economía y Finanzas, más allá de las limitaciones en las capacidades técnicas de las universidades, la mayoría de éstas no cuentan con un marco normativo e institucional para invertir los recursos del canon en un año fiscal (12 meses); las universidades debería contar con herramientas de gestión (planes de investigación y procesos internos para poder ejecutarlos.

Cabe indicar en este mismo informe se analizó la gestión de la transparencia y el acceso a la información en 36 universidades públicas del Perú en relación a la publicación de datos que contienen información de dominio público como por ejemplo: planeamiento y organización; presupuesto; datos generales; proyectos de inversión; contrataciones etc., en el Portal de Transparencia Estándar (PTE), de acuerdo a la Ley de Transparencia 27806: Ley de Transparencia e Información Pública que sirve de base para los estándares en universidades.

2.3.4. Estándares y lineamientos sobre el uso de softwares y metadatos a nivel internacional y nacional

2.3.4.1. Estándares de metadatos a nivel internacional y nacional

Ferrer-Sapena, Peset y Aleixandre-Benavent (2011), señalan que “los proyectos open data pueden contener distintos tipos de información y servicios, es decir, algunos de ellos ofrece únicamente datos, mientras que otros, más avanzados ofrecen además servicios de utilidad para los ciudadanos y empresas. Por tal motivo, se sugiere que esta integración se base en las recomendaciones establecidas por el W3C. Proponen, los siguientes pasos a seguir para su publicación: los datos deben ser publicados en bruto, es decir deben mantener una estructura que permita su uso automatizado (xml (lenguaje de etiquetado extensible), rdf (infraestructura para descripción de recursos) y csv (valores separados por coma); se debe crear un catálogo en línea para que los usuarios puedan conocer qué se ha publicado, recomendando incluir datos que permitan contrastar la calidad de la información aportada; además, los datos deben ser fiables y encontrarse estructurados y documentados; finalmente, se debe enriquecer los datos en formato xhtml (extensible hypertext markup language), a fin de avanzar en el desarrollo de la web semántica y permitir compartir los datos de manera más segura y fiable en todas las aplicaciones y soportes, lo que garantiza la compatibilidad” (p.262 y 263).

Starr, et al.. (2015), en su investigación sobre el esquema de metadatos DataCite para publicaciones y citaciones de investigación, señalan que el consorcio de DataCite, fundado a finales de 2009 ha crecido considerablemente y se ha extendido a nivel

mundial. Tiene tres objetivos fundamentales: establecer un acceso más fácil a los datos de investigación científica en internet; aumentar la aceptación de los datos de la investigación como legítima con contribuciones citables al registro científico; y el apoyo en el archivo de datos que permita la comprobación de los resultados y reutilización para estudios futuros. La clave del servicio de DataCite es el concepto del largo plazo o identificador persistente entre una cadena de caracteres y un recurso y, por tal motivo, ha sido diseñado con los conceptos de flexibilidad y extensibilidad. Los recursos pueden ser archivos, partes de archivos, personas, organizaciones, abstracciones, entre otros. DataCite utiliza identificadores de objetos digitales (DOI).

El esquema de metadatos DataCite es una lista de las propiedades de metadatos básicos elegidos para la identificación precisa y consistente de un recurso para que sea recuperado con éxito, junto con las instrucciones de uso y las recomendadas (p.3) El recurso es típicamente un conjunto de datos. Este esquema de metadatos tiene tres diferentes niveles de obligación en la descripción de las propiedades de los metadatos:

1. Obligatorios (M): propiedades que deben ser proporcionadas;
2. Recomendados (R): propiedades que son opcionales, pero muy recomendables;
3. Opcionales (O): propiedades que no son obligatorias y proporcionan una descripción más rica. DataCite recomienda la inclusión de metadatos obligatorios y recomendados. En las tablas de DataCite, se detalla una descripción de las propiedades obligatorias, junto con sus sub-propiedades que deben ser suministrados con cualquier presentación de metadatos inicial al agente de gestión de DataCite. Starr et al. 2015 (p.5)

Para efectos de la referencia, DataCite prefiere que los nombres DOI se muestren como vinculables, es decir con URL, permanentes.

Para los países de América Latina que forman parte de la Red La Referencia, se han adoptado las normas de Open Aire, entre las cuales se encuentra Open Aire Guidelines for Data Archives. Esta norma define los campos obligatorios y opcionales para el procesamiento e interoperabilidad de los datos y adopta algunas de las normas publicadas por DataCite.

Open Aire Guidelines for Data Archive (OpenAIRE, 2015), ha sido desarrollada con el objetivo de apoyar a los administradores de repositorio de datos para que publiquen sus metadatos de una manera que sea compatible con la infraestructura de OpenAIRE, a fin que todos los contenidos publicados sean interoperables entre las diversas infraestructuras de investigación. Para este fin utiliza el protocolo OAI-PMH.

En esta directiva se describen los criterios que OpenAIRE ha establecido para recolectar los conjuntos de datos. Considera que estos deben cumplir por lo menos con uno de los siguientes: el conjunto de datos es el resultado de un proyecto de investigación financiado, identificado por un identificador de proyecto; y el conjunto de datos está vinculado con una publicación que se encuentre en OpenAIRE.

Por otro lado, este esquema ha adoptado algunos campos del esquema de metadatos de DataCite para recopilar e importar metadatos, sobre todo teniendo en cuenta que DataCite permite realizar la citación mediante el uso de identificadores persistentes tales como: DOI. Así mismo, al igual que DataCite, establece campos: obligatorio (M, que establece que el registro siempre debe estar presente); obligatorio cuando es aplicable (MA, cuando se puede obtener el valor del campo); Opcional (O, cuando se puede proporcionar información complementaria sobre el recurso); y recomendado (R, que recomienda el uso del campo).

Paola Azlirevich (2016) en la video conferencia presentada junto con Silvia Nakano, (Nakano y Azlirevich, 2016), señala que en relación con el trabajo de esquemas de metadatos se trabajó con DataCite 4.0, frente a Open Aire Guidelines for Data Archives y Dublin Core frente a DC Terms. Se analizó Open Aire Guidelines for Literature Repositories, para buscar puntos en común y determinar las diferencias para seleccionar la mejor opción para la región. En setiembre de este año se llegaron a las siguientes conclusiones o propuestas:

1. DataCite es el esquema más adecuado. Sin embargo, se va a esperar el lanzamiento de la versión de Open Aire Guidelines para archivo de datos porque es el esquema que ha elegido la Unión Europea, que sigue el resto del mundo y también porque Open Aire se ha basado en Data Cite hasta hace poco.

2. Los repositorios deberán utilizar otros esquemas de metadatos según estándares más utilizados en el área para poder dar mayor riqueza a sus objetos digitales
3. Es importante registrar los metadatos de relación (dc.relation) que son los que permiten limpiar datos con publicaciones y de financiamiento (Project ID), tanto en los registros de publicaciones como de datos.
4. Pensar en un esquema genérico que permita describir tanto publicaciones como datos, con los mejores criterios de Data Cite y de Open Aire para que los investigadores puedan subir sus datos.

En el Perú el Consejo Nacional de Ciencias, Tecnología e Innovación Tecnología (CONCYTEC), en las Directrices para el procesamiento de información en los Repositorios Institucionales (2016), recomienda utilizar las directrices de DRIVER 2.0 (Digital Repository Infrastructure Vision for European Research), para la recolección de los metadatos y el esquema de metadatos Dublin Core. Adicionalmente, se han adoptado los criterios de DINI Certificate (Directorio de Investigadores) con el fin de fortalecer el trabajo normativo y mantener una concordancia con los estándares internacionales y tecnologías probadas, así como los acuerdos para la interoperabilidad con La Referencia. Así mismo, se incluyen las indicaciones para el ingreso de datos, y los detalles técnicos a considerar para la recolección de metadatos. El uso de algunos campos es repetible y presenta las siguientes características: obligatorio, que exige siempre que se debe ingresar el metadato; recomendado, que sugiere ingresar los metadatos para describir mejor el recurso; y opcional, cuyo uso no es indispensable o no tiene relevancia. Además, en estas directrices se recomienda como plataforma de los repositorios la implementación de DSpace por su flexibilidad e interoperabilidad.

2.3.4.2. Estándares de softwares de datos de investigación a nivel internacional y nacional

Arquero y Marco (2014), señalan que la arquitectura del portal de la Unión Europea “se basa en soluciones de software libre como el sistema de gestión de

contenidos Drupal o el sistema para la gestión de datos CKAN” (p.105). Utiliza tecnologías de la Web Semántica desarrolladas por el World Wide Web Consortium, (W3C) para relacionar los datos como: linked data (datos vinculados), método normalizado que sirve para representar gran variedad de temas, facilitando su relación y su uso (p.114).

El Proyecto Aporta.es de España, (España, Ministerio de Industria, Energía y Turismo, 2015) tiene como objetivo principal incrementar la cobertura y la conectividad de la Red en España. Este proyecto ha desarrollado una guía para publicar datos abiertos de forma rápida mediante CKAN, plataforma de publicación de datos de código abierto. En la primera parte de la guía se establecen las pautas que sirven de guía a la hora de localizar y preparar los datos de su publicación. Explica que para seleccionar los datos que se van a publicar se debe determinar cuál es la información más valiosa a colocar desde el punto de vista de la reutilización para poder dar mayor prioridad a su publicación como: selección de licencia compatible con los principios de datos abiertos que se va a utilizar (Ver tabla 7); y los formatos que se utilizarán, tomando en cuenta que estos deben ser estructurados, es decir que siguen modelos de datos estandarizados; que sean legibles por máquinas; que sean los más demandados entre los investigadores; que sean abiertos a fin que permitan su uso sin restricciones; que permitan publicar metainformación incorporada. (Véase tabla 8). La segunda parte se indica cómo publicar los datos en la Web a través de un catálogo y de la forma más simple posible mediante CKAN.

Tabla 7 Licencias compatibles con los principios de datos abiertos

Licencias compatibles con los principios de datos abiertos		
Nombre	Descripción	Tipo de licencia
Dominio Público	Permiten compartir los datos sin ningún tipo de restricción.	<ul style="list-style-type: none"> • Creative Commons Zero (CC Zero). • Public Domain Dedication and Licence (PDDL 1.0)
Atribuir la fuente original de los datos	Obligan a atribuir la fuente original de los datos	<ul style="list-style-type: none"> • Creative Commons-Reconocimiento (CC BY 4.0) • Open Data Commons Attribution Licence 1.0 (ODC-BY 1.0)
Atribución de la fuente original y compartir los resultados	Requieren atribución de la fuente original y compartir los resultados bajo las mismas condiciones	<ul style="list-style-type: none"> • Creative Commons-Reconocimiento Compartir Igual (CC BY-SA 4.0) • Open Data Commons Open Database License (ODbL 1.0). • Open Data Licence Attribution

Fuente. Elaboración propia

Tabla 8 Formatos frecuentemente utilizados para la apertura de la información

Ejemplo de formatos utilizados para datos abiertos	
Formato	Función
CSV(TSV)	Para representar datos en forma de tabla, donde la columna se separan por comas (o tabulaciones) y las filas por salto de línea.
ODS/XLS	Formatos (abierto en el caso de ODS y propietario de Microsoft en el caso de KLF). De hoja de cálculo que muestran la información en celdas organizadas en filas y columnas. Cada celda contiene datos o una fórmula con referencias relativas, absolutas o mixtas a otras celdas.
GML	Vocabulario XML para el modelaje, transporte y almacenamiento de información geográfica.
KML	Vocabulario XML para representar datos geográficos en tres dimensiones. Suele distribuirse comprimidos en formato KMZ.
JSON	Notación muy popular para el intercambio de datos basada en una simplificación de la notación de Java Script.
RDF	Modelo de representación para la descripción de recursos siguiendo los fundamentos de la Web Semántica.
RSS	Vocabulario XML para facilitar la distribución de contenidos y actualizaciones.
SHP	Formato propietario de datos espaciales que se ha convertido en un estándar, de-facto y sirve para almacenar tanto la localización de los elementos geográficos como los atributos asociados a ellos.
WMS	Formato de servicio Web para producir mapas de datos referenciados espacialmente, de forma dinámica a partir de información geográfica.
XML	Lenguaje general de etiquetado que define un conjunto de reglas para la codificación de documentos a través de modelos específicos legibles y procesables.

Fuente. España. Ministerio de Industria, e. y T.E (2015)

2.3.4.2.1. Software CKAN

A fin de establecer una gestión adecuada de los datos para su reutilización, la guía recomienda establecer una recopilación de los mismos con un solo punto de acceso en un catálogo de datos, lo cual permitirá una organización más eficaz y un acceso más fácil para los investigadores.

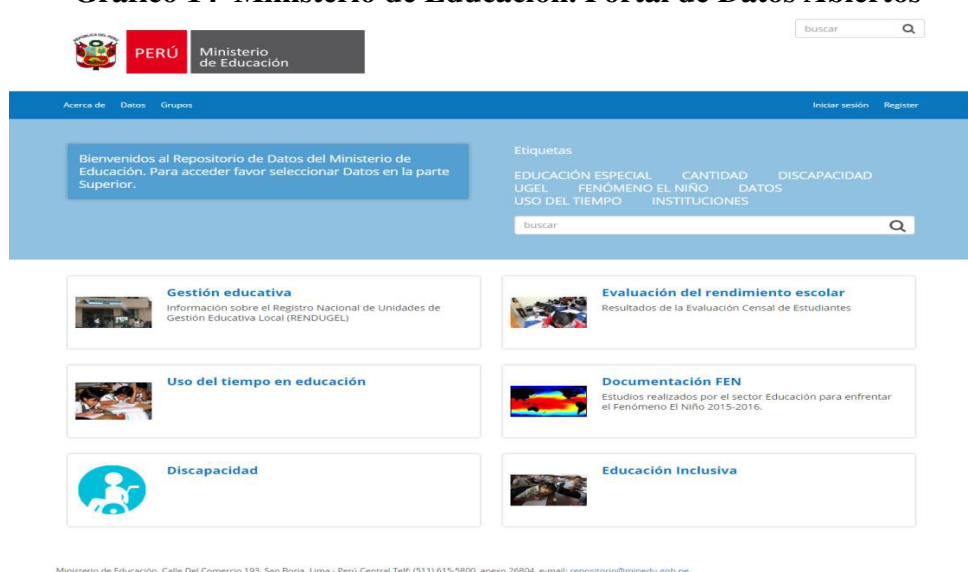
CKAN (Comprehensive Knowledge Archive Network) es una aplicación web de código abierto que está concebida para la publicación de catálogos de datos abiertos. Esta solución está dirigida a proveedores de datos (gobiernos nacionales, regionales, locales, empresas y organización) que deseen proporcionar datos abiertos y reutilizables. Actualmente es el software de repositorio de datos de código abierto más utilizado por gobiernos que cuentan con diversos portales de datos abiertos como los portales de la Comunidad Europea, Reino Unido y Australia. Fue creada por la Open Knowledge Foundation (OKF). De acuerdo con lo señalado en la guía Aporta.es “se podría comparar con sistema de gestión de contenidos (CMS) pero orientado a la publicación de datos. Permite, publicar, almacenar y distribuir colecciones de datos con muy pocos recursos y una vez publicados, los usuarios podrán explorarlos e incluso pre-visualizarlos a través de mapas, gráficas y tablas”. (España. Ministerio de Industria, Energía y Turismo (2015), (p.7).

2.3.4.2.2. Software DKAN

Otro software de código abierto muy utilizado, es el DKAN. Es un software muy parecido al CKAN, reescrito en PHO, utiliza el sistema de control de contenidos de Drupal. Está construido sobre características tradicionales de un portal de datos, ofrece una navegación muy amigable con diferentes facetas de búsqueda a través de miles de recursos y utiliza estándares internacionales de Open Data, para la publicación y gestión de los datos.

En el Perú el DKAN, es utilizado por la mayoría de instituciones que cuentan con portales de datos abiertos como es el caso del CONCYTEC, el Ministerio de Educación, (Ver gráfico 14), Ministerio de Salud, (Ver Gráfico15).la Presidencia del Consejo de Ministros (PCM), entre otros.

Gráfico 14 Ministerio de Educación. Portal de Datos Abiertos



Fuente. Perú. Ministerio de Educación (2016).

Gráfico 15 Ministerio de Salud. Portal de Datos Abiertos



Fuente. Perú. Ministerio de salud (2016).

2.3.5. Investigación en universidades: Facultades de Ciencias de la Salud

2.3.5.1. Investigación en las universidades en Perú

En la monografía presentada por el Ministerio Federal Alemán para la Educación y la Investigación y la Oficina Internacional del BMBF alemán, *Mapa de Investigación, Perú, Perú* (Kiwitt-López, 2011), se señalan que los mapas de investigación sirven para detallar y describir las estructuras, estrategias y actores de un país en relación a la investigación, tecnología, innovación y educación. El objetivo de estos mapas es facilitar a científicos, instituciones científicas y representantes de organizaciones de investigación y de fomento de la investigación en Alemania la búsqueda de socios cooperadores, temas y programas en los países protagonistas correspondientes. Sirven de base para una correcta planificación y diseño de futuras colaboraciones bilaterales, tanto a nivel científico como político. En este trabajo se menciona que los desarrollos políticos y económicos de los países tienen una gran influencia en el ámbito de la investigación y la innovación. En el Perú en los años 70, con la fundación del Consejo Nacional de Investigaciones (CONI), así como otras instituciones de investigación y diferentes iniciativas legales, se tomaron las primeras medidas para la estructuración del sistema nacional de investigación y la financiación de la investigación, la tecnología y el desarrollo. Sin embargo, no fue sino hasta el gobierno de Alejandro Toledo (2001-2006), que se realizaron los primeros pasos para reforzar el sistema de investigación. Para este fin se promulgaron diferentes leyes, se desarrollaron planes de investigación y se diseñó un programa de fomento con medios internacionales.

La competencia del estado en ciencia y tecnología se encuentra principalmente a nivel nacional. El organismo más importante en el Sistema Nacional de Ciencia Tecnología e Innovación Tecnológica (SYNACIT) es el Consejo Nacional de Ciencia, Tecnología e innovación Tecnológica CONCYTEC. Este organismo dirige, coordina, evalúa y promueve el sistema de investigación peruano. Por otro lado, muchos institutos de investigación y otras instituciones relacionadas con la investigación, la tecnología y la innovación están subordinadas a ministerios.

Rubio, (2009) en la tesis *La estructura organizacional en Centros de Investigación y Desarrollo e Innovación; una aproximación a la experiencia internacional* señala que “el conocimiento es el elemento fundamental para el desarrollo y competitividad de los países y, por tanto, del bienestar social de los mismos; este se genera principalmente en los centros de investigación, desarrollo e innovación, CI+D+I, tanto de universidades como independientes, públicos o privados.

De acuerdo con el artículo publicado por Deis, (2001), *La investigación básica en América Latina*, se señala que “sin investigación básica, no hay desarrollo tecnológico, asimismo manifiesta que el apoyo a la investigación científica básica y aplicada es una decisión política y obedece a una conducta universitaria y en tanto no exista la decisión política de apoyarla, las universidades serán indiferentes a la formación de jóvenes científicos.

2.3.5.2. Facultades de Ciencias de la Salud

En el artículo publicado por Salaverry, (2006), *El inicio de la educación médica moderna en el Perú. La creación de la Facultad de Medicina de San Fernando*, se realiza una reseña sobre la enseñanza de la medicina en el Perú. Este artículo describe que esta “se remonta al siglo XVI con la creación de las cátedras clásicas de medicina en la Universidad San Marcos (p.122). Sin embargo, recién en 1855 con el Gobierno de Castilla, se realizó una reforma educativa a nivel nacional en la cual se presenta una propuesta para el desarrollo de la Facultad con varios puntos importantes, entre ellos, el plan de estudios en el cual se establece el régimen lectivo semestral en lugar del anual que estaba desde sus inicios; así mismo, exige como condición para la matrícula de los alumnos que este cuente con el bachillerato en Filosofía y Matemáticas. Finalmente hace énfasis en la enseñanza práctica en los hospitales desde el segundo año de la Facultad. Con todas estas medidas se establecen las bases de los estudios de las Facultades de Medicina en el Perú.

Risco de Domínguez, (2014) *Diseño e implementación de un currículo por competencias para la formación de médicos* describe el “proceso de creación de una escuela de Medicina en el Perú, basada en competencias (EBC). Este modelo, se

comienza a implementar a partir de los años 90 en varias escuelas de Medicina y actualmente es el modelo más recomendado por los principales organismos rectores y evaluadores de la educación médica a nivel mundial”. Así mismo, señala que en el Perú el EBC fue incorporada como un estándar de calidad de la educación médica por la Comisión de Acreditación de Facultades o Escuelas de Medicina Humana (CAFME) en el año 1999, en la cual se establece la acreditación obligatoria de las facultades de o escuelas de Medicina y suspendía la creación de nuevas facultades o escuelas.. En el año 2011 se establecieron los estándares mínimos para la evaluación previa por CAFME en relación a la creación de nuevas escuelas de Medicina que comprenden: la organización; los procesos; los recursos y servicios en lo que se refiere al currículo y se prescribe el currículo por competencias en el estándar F. CAFME. Se considera la inclusión de la EBC como estándar necesario para asegurar un adecuado y correcto desempeño profesional de los estudiantes de Medicina”. Risco de Domínguez (p.573).

La Asociación Peruana de Facultades de Medicina- ASPEFAM (Asociación Peruana de Facultades de Medicina (2016) presenta en su página web una relación de las facultades de Medicina en el Perú. De acuerdo con esta relación existen 31 facultades en distintas ciudades del Perú. (Ver anexo 2).

2.3.5.2.1. Ciencias de la Salud en la Universidad Nacional Mayor de San Marcos (UNMSM).

Según el artículo publicado por Pacheco-Romero, Huerta, Galarza y Ramos-Casillo (2012), se describe que el Vicerrectorado de Investigación de la Universidad Nacional Mayor de San Marcos, fue creado en junio de 2005, mediante la RR N° 03316-R-05, debido a la decisión estratégica de la universidad de convertirse en la universidad líder en producción de conocimientos del país. El Vicerrectorado se creó con el fin de promover, evaluar y posibilitar las actividades de investigación en la Universidad. Por tal motivo, se empezó con el ordenamiento de la investigación, el monitoreo, la creación de puntajes del investigador de acuerdo con la producción y publicación científica, el estímulo para los investigadores de las diversas facultades y

en particular haciendo énfasis en la Facultad de Medicina. Para este fin se organizaron talleres de capacitación y perfeccionamiento de los concursos de investigación con apoyo económico, el establecimiento de presentar informes técnicos y económicos de acuerdo a plazos establecidos.

El artículo mencionado trata de la organización de talleres dirigidos a capacitación en investigación en la Facultad de Medicina de la UNMSM en el último quinquenio (2004-2009). En el año 2009 con el apoyo de la Organización Panamericana de la Salud, se realizó un Taller de Análisis Estratégico sobre Investigación en la Facultad de Medicina de San Fernando, con el fin de señalar la necesidad de conocer la producción de cada investigador de los diversos Institutos de Investigación de la Facultad: Instituto de Biología Andina, Instituto de Cirugía Experimental, Instituto de Ética en Salud, Instituto de Investigaciones Clínicas, Instituto de Medicina Tropical e Instituto de Patología, y el Centro de Investigación en Bioquímica y nutrición de manera que los resultados permitieran tomar las medidas de perfeccionamiento en investigación. Este taller, se realizó debido a las inquietudes presentadas por los docentes, debido a que no tenían tiempo suficiente para investigar. El estudio demostró que 14 (25.9%) docentes tuvieron que disminuir su carga docente de manera de poder investigar.

Salaverry, (2006), señala que la creación de la Facultad de Medicina de San Fernando en San Marcos se llevó a cabo el 9 de setiembre de 1856, en el marco de la reforma educativa emprendida por el gobierno del mariscal Castilla. El área de Ciencias de la Salud de la UNMSM está conformada por 6 facultades: la Facultad de Medicina Humana, cuenta con 6 Centros de Investigación: Facultad de Medicina Humana, cuenta con la Unidad de Investigación de la Facultad de Medicina Humana que coordina las actividades de cuatro institutos y un centro de investigación, con la finalidad de contribuir y aportar conocimientos y soluciones a los problemas de salud de la población peruana; Facultad de Farmacia y Bioquímica, cuenta con la Unidad de Investigación de la Facultad de Farmacia y Bioquímica. Coordina las actividades de dos institutos y un centro de investigación. Los temas que investiga son: plantas medicinales, medicamentos, ecología y toxicología, calidad de los alimentos, bioquímica, microbiología y biotecnología; Facultad de Odontología cuenta con el

Instituto de Investigación Estomatológica que se enfoca en estudios clínico-patológicos, socio-epidemiológicos y de biopatología experimental de la cavidad oral con énfasis en el aspecto educacional; Facultad de Medicina Veterinaria cuenta con el Instituto Veterinario de Investigaciones Tropicales de altura (IVITA). Desarrolla proyectos de investigación en el campo pecuario, orientados principalmente a la producción y productividad y salud animal, así como al manejo de recursos naturales; Facultad de Psicología cuenta con el Instituto de Investigaciones Psicológicas. Sus líneas de acción son: Psicología clínica y de la salud, Psicología educacional y Psicologías social organizacional; y la Facultad de Ciencias Biológicas, cuenta con el Instituto de Investigación de Ciencias biológicas “Antonio Raimondi” (ICVAR). Desarrolla líneas, programas y proyectos de investigación de ecología, microbiología, control biológico, biotecnología, biología celular, genética, parasitología, conservación, manejo y producción de recursos biológicos. Prioriza estudios de conservación y utilización de la biodiversidad tanto de organismos como de ecosistemas.

Por otro lado, cabe indicar que actualmente, según Scimago Institutions Ranking (SIR) 2015, la Universidad Cayetano Heredia tiene 1,228 papers académicos indexados a la base de datos Scopus, superando a la Universidad Nacional Mayor de San Marcos, que tiene 892 y a la Pontificia Universidad Católica del Perú que cuenta con 644 papers. En este sentido, la Universidad Cayetano Heredia se convierte actualmente en la pionera en investigación en el Perú y la cuarta en Sudamérica.

2.3.5.2.2. Facultad de Ciencias de la Salud en la Universidad Peruana de Ciencias Aplicadas UPC.

Risco de Domínguez, (2014) señala que bajo los estándares de calidad establecidos por la Comisión de Acreditación de Facultades o Escuelas de Medicina Humana (CAFME) en el año 2001, en el cual se incluye como estándar para la acreditación de las facultades de medicina el establecimiento de un currículo basado en competencias (EBC), la Universidad Peruana de Ciencias Aplicadas (UPC), decidió crear en el año 2005 una nueva escuela de Medicina (EM-UPC). La elaboración del proyecto tomó 14 meses y fue presentado a CAFME en febrero de 2006 y aprobado

por Resolución del 25 de octubre de 2006. La EM-UPC inició su funcionamiento en marzo de 2007.

Risco de Domínguez (2014) señala que en “el proceso de diseño del currículo se llevó a cabo basado en los lineamientos del modelo educativo de la UPC, en el análisis de la literatura sobre las principales tendencias internacionales en educación médica y la visita a importantes escuelas de Medicina de Estados Unidos, Europa y Latinoamérica. Luego de haber realizado el análisis se diseñó el currículum incluyendo: los resultados que se espera alcanzar, los contenidos, las competencias, las estrategias, la metodología, la evaluación y los recursos a emplear, articulando todos los componentes, para el logro de los resultados, que responden a las necesidades de la población y de los sistemas de salud. En este contexto, se analizó también las características de la educación médica en el Perú para el año 2005, (Ver tabla 9).

Tabla 9 Características de la educación Médica en el Perú para el 2005

Temas	Características
La formación médica	Es teórica, técnica, basada en contenidos. “Se forman médicos que el país ya no necesita”.
Los egresados	No están preparados para desempeñarse en el SERUMS
Los docentes	Especialistas, no tienen formación docente.
El currículo	Recargados de contenidos, gran importancia de la clase magistral y del rol del profesor como trasmisor de conocimientos.
Recomendaciones: Perfil del nuevo médico	Médico general. Competencias personales desarrolladas: innovador, con pensamiento crítico, ética y que sea modelo de trato humano y de respeto a la privacidad de los pacientes. Competente para curar y prevenir enfermedades. Buena formación científica, hace investigación y publica.
Recomendaciones: La nueva escuela de Medicina	Innovadora, aplicar nuevas metodologías de aprendizaje y nuevas tecnologías. Aplicar aprendizaje basado en problemas., Hacer docencia en hospitales de niveles I,II y III. Actividades educativas en la comunidad. Debe tener un hospital universitario

Fuente: Risco, Domínguez G. (2014)

Manifiesta e que el perfil profesional está muy ligado al curriculum y es la base del proceso del diseño curricular, vinculándolo en el contexto en el cual se desarrolla. El perfil profesional del médico UPC consta de diez competencias: cinco generales o de formación de la persona y cinco específicas, propias de la profesión. (véase tabla N° 13). Los cursos y el currículo se constituyen con base en las competencias del perfil profesional y sus criterios de logro.

Para la autora, la competencia es mucho más que un listado de conocimientos, habilidades, y actitudes que se dan simultáneamente y que el alumno debe memorizar; es un aprendizaje complejo que requiere la integración de estos tres elementos en forma integral de modo que se conviertan en un comportamiento integral y esto se logra cuando el estudiante participa en forma activa en experiencias de aprendizaje

adecuadas. Por tal motivo, señala que las estrategias que se emplearon en el diseño del currículo de la EM-UPC son el colocar al estudiante en el centro del proceso de enseñanza-aprendizaje; el empleo de la metodología de aprendizaje basado en problemas (AABP, TBL, simulación etc.) y la integración curricular.

La carrera de Medicina de la UPC, se creó en el año 2005 como Escuela de Medicina y empezó a funcionar en el año 2007. En el año 2008 se convirtió en la Facultad de Ciencias de la Salud conformada por cuatro carreras o programas. La carrera de Medicina, así como la carrera de Nutrición y dietética tienen como líneas de investigación para sus estudiantes las siguientes: Biomateriales; biología sintética y bionanotecnologías; Bioquímica, biología molecular y celular; Educación en Ciencias de la Salud; Epidemiología clínica y salud basada en evidencias: Nutrición, deporte y actividad física; Salud Pública y gestión en salud. Las carreras de Odontología y Terapia Física tienen las mismas líneas de investigación pero dirigidas hacia la odontología y terapia física. Cabe indicar, que a partir del año 2016 la carrera de Odontología ofrecerá segunda especializaciones como: segunda especialidad en endodoncia; segunda especialidad en ortodoncia y ortopedia; segunda especialidad en odontología pediátrica; segunda especialidad en periodoncia e implantes; y segunda especialidad en rehabilitación oral.

Desde la creación de la Facultad de Ciencias de la Salud, en el año 2007, la UPC se planteó como una de sus líneas estratégicas, realizar investigación científica y publicar trabajos de investigación sobre temas de importancia para el análisis y solución de los problemas de salud de la población peruana. Por tal motivo, la UPC firmó un Convenio de afiliación entre el Instituto de Investigación Nutricional (INN) y contratar a un investigador a tiempo completo. Durante ese año la UPC, compitió con 300 proyectos para lograr el otorgamiento de un Grant del Fondo para la Innovación, la Ciencia y la Tecnología (FINCyT). El proyecto presentado “Diagnóstico de virus respiratorios y gérmenes atípicos en pacientes inmunodeprimidos mediante PCR multiplex” que a la fecha está concluido con 1500 pacientes y en fase de preparación para ser publicado en una revista de investigación.

A enero de 2011 el laboratorio de Investigación se encontraba inmerso en tres proyectos que cuentan con financiamiento para su ejecución y dos proyectos que se van a presentar para obtener financiación. Lo más importante es que los alumnos participan en estos proyectos, de esta manera aprender a través del método científico.

En el año 2012 se crea la Dirección de Investigación en la UPC. Esta Dirección junto con el Vicerrectorado Académico y de Investigación y la Facultad de Ciencias de la Salud son los que marcan las pautas relacionadas con la investigación.

En el 2013, se creó el Centro de Investigación en Ciencias de la Salud de la UPC, se cuenta con un grupo de investigadores altamente calificados, investigadores visitantes y personal técnico especializado. Este Centro es multidisciplinario, lo que permite realizar investigaciones en todos los ámbitos del conocimiento y contribuir al avance científico-tecnológico mediante el desarrollo de nuevas estrategias de competitividad frente a las demandas de la sociedad. Sus laboratorios constan de 5 áreas: Biología Molecular, Biología Celular, Microbiología, Bióquímica y Proteómica. Se destacan en este Centro las líneas de investigación de identificación y caracterización de virus y bacterias causantes de infecciones respiratorias agudas (IRAS), enfermedades febriles, gastroenteritis aguda, diagnóstico y prevención de cáncer de cuello uterino y cáncer de estómago; así como estudios sobre principios activos de plantas con propiedades antivirales y antibacterianas, entre otros. Las áreas que más investigan en el centro de investigación son las áreas de Biología Celular y Molecular de las enfermedades Infecciosas y Enfermedades Respiratorias.

Dentro de los fines más importantes de este Centro están:

- Desarrollar la investigación e innovación en ciencias biomédicas en la Facultad;
- Potenciar las carreras de ciencias de la salud, incorporando investigadores de prestigio, quienes participarán en la docencia de diferentes niveles y propiciando la incorporación de alumnos a los proyectos de investigación.
- Desarrollar productos biomédicos y biotecnológicos.
- Generar la publicación de artículos científicos referentes a los resultados de los proyectos de investigación efectuados por los investigadores del Centro.

- Atraer fondos públicos y privados para financiar proyectos de investigación, desarrollo tecnológico e innovación.
- Formar nexos multidisciplinarios al interno y externo de la Universidad.
- Brindar infraestructura, equipamiento y recursos básicos para promover investigaciones biomédicas entre docentes y alumnos.
- Promover convenios de colaboración científica con investigadores nacionales y extranjeros.

En este Centro participan profesores investigadores nombrados por el Director de Investigación de la Universidad y el Decano de la Facultad, quienes lideran el Centro, profesores investigadores invitados y financiados por el FONDECYT, profesores, tesis y alumnos de la Facultad quienes colaboran y dan soporte en las investigaciones del Centro y asistentes científicos, de otras facultades y universidades nacionales y extranjeras que colaboran y dan soporte en las investigaciones del Centro.

Es importante resaltar que la investigación es inherente al estudio de la Medicina desde tiempos remotos, así como también la publicación de sus investigaciones. Por tal motivo, la Facultad de Ciencias de la Salud de la UPC ha contribuido enormemente en el incremento de la producción de artículos científicos de la Universidad que se encuentran en la base de datos Scopus, tal como se demuestra en el gráfico adjunto. Se puede decir que en el Perú no hay otra universidad que en tan poco tiempo haya crecido en el número de artículos científicos. Por este motivo, se debe aprovechar poder gestionar los datos de investigación que han dado como resultado la publicación de artículos científicos como una oportunidad de desarrollo y para ello se necesita saber la percepción de los docentes investigadores de la Facultad de Ciencias de la Salud en la gestión de los datos a fin de poderlos gestionar en un repositorio de datos.

2.4. Cuadro comparativo de Modelos del ciclo de vida de los datos y el modelo propuesto

Los modelos que han servido como guía para realizar la propuesta del modelo de gestión de los datos basados en el ciclo de vida, son los proporcionados por el Digital Curation Center (DCC) y UK Data Archive. (p.41). Sobre la base de estos modelos se

ha elaborado el cuadro comparativo del análisis del ciclo de vida de los datos (modelos existentes); y el modelo propuesto (Ver tabla 10).

El esquema general de los modelos de ciclo de vida, se inicia con la creación y el almacenamiento de los mismos hasta el momento en que este se convierte en obsoleto y es desechado. En el modelo propuesto, lo primero que debe estar listo y aprobado por la institución es el establecimiento de un plan de gestión de datos de investigación en el cual se debe tomar en cuenta elementos como: los investigadores (sus necesidades y requerimientos); lineamientos y políticas para compartir que estén de acuerdo a las directrices de la institución; tipos de datos que se crearán; formatos que se utilizarán; descripción del conjunto de datos; normas y metadatos; como serán citados los datos; privacidad, seguridad y derechos de autor; metodología de intercambio de datos; metodología de preservación y conservación; roles y responsabilidades del investigador, y gestor de la información enfocado en el acompañamiento al investigador. Lo que proponemos es que el gestor de la información debe realizar el acompañamiento al investigador, durante todo el ciclo de vida de los datos, con el propósito de que comparta y reutilice datos de investigación desde su creación hasta que dejen de ser útiles.

La contribución del modelo propuesto es la integración, al de ciclo de vida de los datos, de un proceso crítico como es la difusión, que definimos en dos líneas de acción. La primera se enfoca en las actividades de difusión sobre los beneficios del compartir y reutilizar los datos de investigación, bajo los criterios de eficiencia, calidad y economía en el desarrollo de la investigación; y la segunda se enfoca en las actividades de captación y reclutamiento de docentes investigadores del Área de Ciencias de la Salud, bajo los criterios de aumento del número de investigaciones y generación de nuevos conocimientos. Se debe tener presente, que las actividades de difusión deben ser apoyadas por un plan de actividades de inducción y publicidad.

Tabla 10 Análisis Comparativo de Modelos del Ciclo de vida de los datos implementados con modelo propuesto

Digital Curation		UK Data Archive		UNMSM-UPC	
Acciones secuenciales		Etapas para el ciclo de vida		Etapas	
Crear o recibir	Proceso de generación o recepción de los datos. Asegurarse de que los datos son recogidos en un formato adecuado y descritos con metadatos apropiados.	Creación de datos:	Diseño de la investigación, creación de planes de gestión de datos, políticas de data sharing, localización de los datos, recolección y creación de metadatos.	Crear o recibir	Generación del dato por el investigador. Proceso de localización y recolección de los datos generados por los investigadores. Recepción de los datos en formatos adecuados y descripción general de los metadatos.
Preservar	Acciones que se deben realizar para la preservación a largo plazo: acciones de corrección de errores, de limpieza de datos, validación de formatos, copias de seguridad etc.	Preservación de los datos:	Migración de los datos al formato recomendado, creación de backups, creación de metadatos y documentación, archivo de datos.	Preservar	Aseguramiento del dato de investigación a corto, mediano y largo plazo. Utilidad del dato a fin de poder ser desechados, trasladados a otro archivo, depósito, centro de datos o destruidos de acuerdo a la política establecida.
Acceder, usar y reutilizar	Incluir metadatos, realizar indexaciones, controlar el acceso a los datos y establecer licencias y permisos.	Dar acceso a los datos:	Distribuir, compartir y controlar el acceso a los datos, como establecer licencias y promocionar los datos.	Acceder	Acercarse a datos originales, fiables, que se encuentren organizados y documentados y que cumplan con estándares de esquemas de metadatos.
Evaluar y seleccionar	De acuerdo a las políticas establecidas se realizar una evaluación de los datos para seleccionar los que van a ser preservados a largo, mediano o corto plazo.	Análisis de los datos:	Interpretación y preservación de los datos para su preservación.	Evaluar: interpretar; analizar; y seleccionar	De acuerdo a las políticas establecidas en el plan de gestión de datos, se debe interpretar los datos para analizarlos, seleccionarlos y darles valor.
		Procesamiento de datos:	Introducción, digitalización, transcripción, traducción, validación, limpieza, descripción, almacenamiento y protección de datos (en los casos que deba asegurarse el anonimato)		
		Reutilizar los datos:	Realizar un seguimiento de la investigación, retomar nuevos resultados, enseñar y aprender de ellos"	Reutilizar	Validación de los datos de investigación almacenados para que puedan ser utilizados en una nueva investigación de acuerdo a las licencias establecidas para los datos abiertos, así como a los estándares de metadatos y softwares.
Desechar	Los datos pueden ser desechados o trasladados a otro archivo, depósito, centro de datos o destruidos de acuerdo a la política establecida.				
Traspasar	Transferencia de datos para que sean reutilizados por otros investigadores.				
Almacenar	Establecer un proceso de almacenamiento de datos sencillo y proporcionar apoyo y orientación siempre que sea posible. Automatizar los procesos y definir quién es el responsable de garantizar la calidad de los datos: el investigador, el gestor de información, etc.			Almacenar	Concentración de datos para ser compartidos, reutilizados y preservados en una plataforma que permita la gestión para su recepción y envío.
Transformar	Creación de nuevos datos a partir de los originales a un formato diferente, para evitar la obsolescencia, creación de nuevos resultados derivados e iniciar el nuevo ciclo de vida de los datos.			Transformar	Creación de nuevos datos a partir de los originales a un formato diferente. Generación de nuevos resultados que dan inicio a un nuevo ciclo de vida de los datos.
				Difundir	Facilitar el conocimiento de los beneficios de compartir y reutilizar los datos de investigación. Captación y reclutamiento de docentes investigadores del Área de Ciencias de la Salud.

CAPÍTULO III: MODELO PROPUESTO PARA LA GESTIÓN DE DATOS DE INVESTIGACIÓN EN BASE AL CICLO DE VIDA DE LOS DATOS

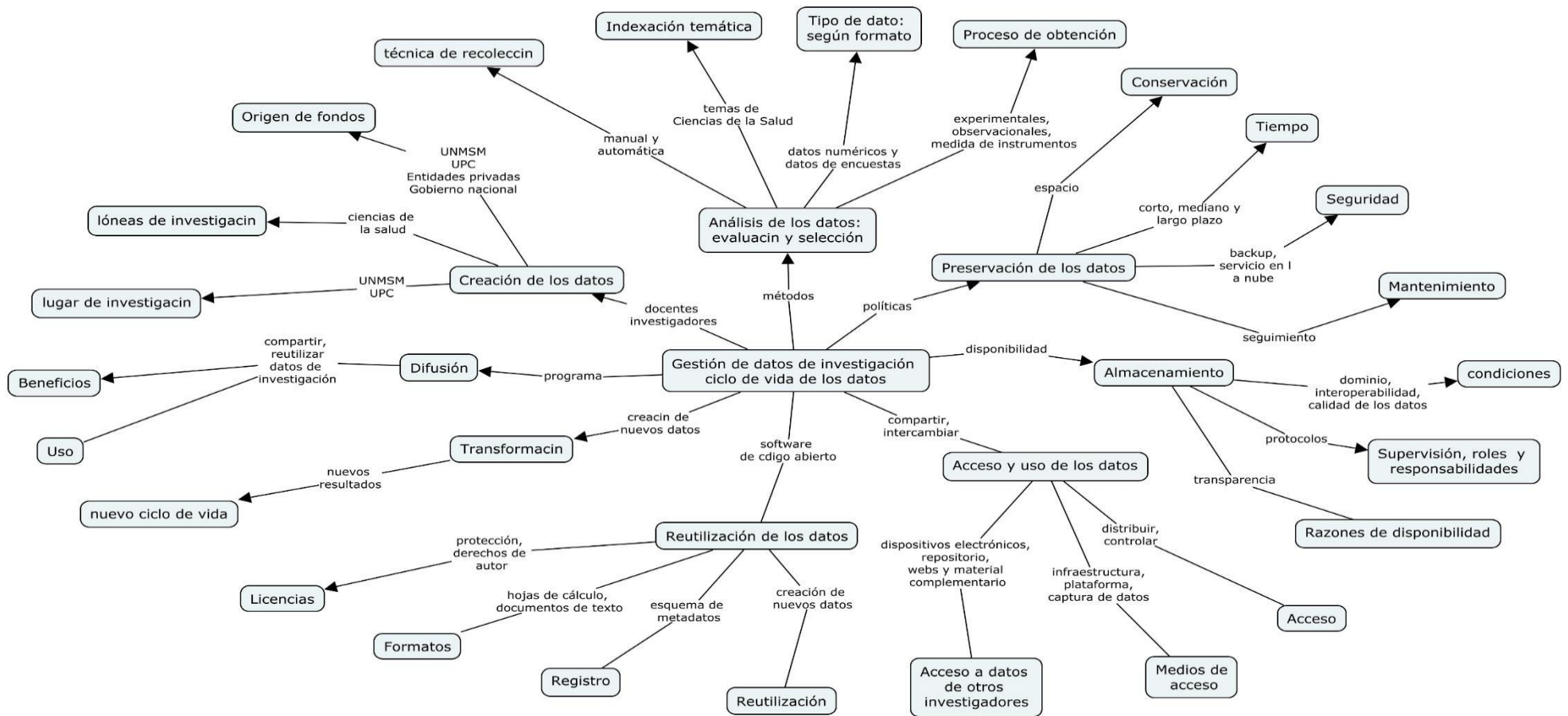
Para incrementar el número de investigaciones, mejorar su calidad y generar nuevos conocimientos, se requiere que los docentes investigadores de las universidades, tanto públicas como privadas, que se desarrollan en el campo de las Ciencias de la Salud, compartan, intercambien y reutilicen sus datos de investigación. Para lograr este propósito es necesario que la producción de datos de investigación se concentre en una plataforma informática (un blog, un repositorio u otro) y sean gestionados de manera eficiente. Para ello, se requiere estructurar un modelo que reúna los elementos necesarios para facilitar el uso e intercambio de datos de investigación, la interacción entre investigadores, organizar los datos de investigación que permita ahorrar tiempo y esfuerzo, que el manejo de los datos de investigación sea transparente y asegure su preservación y que proteja los derechos de autor y la propiedad intelectual. A partir de esta idea central y con la ayuda de la revisión bibliográfica, obtuvimos el marco teórico suficiente que sostiene nuestra propuesta de un modelo para la gestión de datos de investigación basado en el ciclo de vida de los datos.

3.1. Mapa conceptual de la gestión de datos de investigación en base al ciclo de vida de los datos.

Como primer paso para el diseño del modelo propuesto, elaboramos un mapa conceptual, (ver Gráfico 16), considerando todos los elementos importantes que intervienen en la gestión de los datos de investigación basada en el ciclo de vida de los datos. Se ha tenido presente: las políticas respecto a las licencias de protección de los derechos de autor y propiedad intelectual; la utilización de estándares de metadatos y software (DKAN), que permitan registrar los datos de investigación adecuadamente; elección de los formatos de almacenamiento idóneo; selección de las técnicas de recolección de datos; y definición del tipo de datos de investigación a preservar. En lo que se refiere a la operatividad del modelo, este deberá contar con una plataforma informática, que cuente con protocolos interoperables y con parámetros específicos

sobre: espacio, tipo de almacenamiento, disponibilidad, preservación, transparencia, validación y uso de resultados. Otro aspecto importante es la elaboración del plan de gestión de datos y la calificación de un gestor de información especializado o experto, que acompañe al investigador en la elección de estándares, licencias de acceso abierto y capacitación en el uso de formatos con lineamientos y políticas claras.

Gráfico 16 Mapa conceptual de la gestión de datos de investigación en base al ciclo de vida de los datos



Fuente. Elaboración propia

3.2. Modelo Propuesto

A partir de esta concepción, se plantea el diseño del modelo para la gestión de datos de investigación basado en el ciclo de vida de los datos, (Ver Gráfico 17).

El modelo propuesto es una aproximación teórica de hechos reales sobre el uso, la organización y el almacenamiento de los datos de investigación, el modelo contribuye a explicar estos hechos y como se relacionan a partir de la gestión de los datos de investigación durante su ciclo de vida.

Gráfico 17 Modelo Propuesto para la Gestión de Datos de Investigación basado en el Ciclo de Vida de los Datos



Fuente. Elaboración propia

De los resultados del estudio se desprende que un elemento crítico para impulsar el interés de los investigadores por la gestión de datos de investigación en acceso abierto es la difusión de los beneficios de compartir y reutilizar los datos de investigación. (Ver Tabla 11).

Tabla 11 Etapas del modelo propuesto basado en el ciclo de vida de los datos de investigación

UNMSM–UPC		
Etapas	Concepto	Indicador
Crear o recibir	Generación del dato por el investigador. Proceso de localización y recolección de los datos generados por los investigadores. Recepción de los datos en formatos adecuados y descripción general de los metadatos.	Disponibilidad de datos de investigación
Evaluar: interpretar; analizar; y seleccionar	De acuerdo a las políticas establecidas en el plan de gestión de datos, se debe interpretar los datos para analizarlos, seleccionarlos y darles valor.	Tipo de datos de investigación
Preservar	Aseguramiento del dato de investigación a corto, mediano y largo plazo. Utilidad del dato a fin de poder ser desechados, trasladados a otro archivo, depósito, centro de datos o destruidos de acuerdo a la política establecida.	Duración de la conservación de los datos de investigación
Almacenar	Concentración de datos para ser compartidos, reutilizados y preservados en una plataforma que permita la gestión para su recepción y envío.	Condiciones de envío de datos de investigación
Acceder	Acercarse a datos originales, fiables, que se encuentren organizados y documentados y que cumplan con estándares de esquemas de metadatos.	Datos generados utilizados por otros investigadores
Reutilizar	Validación de los datos de investigación almacenados para que puedan ser utilizados en una nueva investigación de acuerdo a las licencias establecidas para los datos abiertos, así como a los estándares de metadatos y softwares.	Conservación de datos de investigación
Transformar	Creación de nuevos datos a partir de los originales a un formato diferente. Generación de nuevos resultados que dan inicio a un nuevo ciclo de vida de los datos.	Disponibilidad de datos de investigación
Difundir	Facilitar el conocimiento de los beneficios de compartir y reutilizar los datos de investigación. Captación y reclutamiento de docentes investigadores del Área de Ciencias de la Salud.	Investigadores que comparten y reutilizan datos de investigación

Fuente. Elaboración propia

3.3. Etapas del ciclo de vida de los datos del modelo propuesto (ver Tabla 11)

3.3.1. Creación de los datos de investigación

El ciclo de vida de los datos se inicia desde el momento en que estos son creados e identificados por los investigadores. En el modelo propuesto la creación de los datos se inicia con la generación de un nuevo dato a gestionar de acuerdo a las líneas de investigación que más destacan o temas más desarrollados y a las necesidades y requerimientos de los investigadores como: Educación en Salud y Epidemiología. En esta etapa se deben recoger y asignar los metadatos descriptivos a los datos proporcionados por los investigadores de acuerdo a estándares internacionales para asegurar una descripción adecuada, previa revisión del plan de gestión de datos de la institución. Los metadatos descriptivos son los que ayudan a describir e identificar un recurso de información para que sean recuperados con éxito. Starr et al. 2015 (p.5)

3.3.2. Evaluación, análisis y selección de los datos.

De acuerdo a la política establecida en el plan de gestión de datos se debe realizar una evaluación de los datos para saber cuáles van a ser preservados y cuáles desechados, tomando como base el interés académico y administrativo de los mismos a fin de hacerlos accesibles y reutilizables por otros investigadores, ayudándolos en la localización de los mismos para que puedan ser relevantes para sus posteriores investigaciones. En esta etapa se debe procesar y analizar si los datos tratan del mismo tema u otro distinto para que a partir de ellos se pueda construir información y posteriormente el conocimiento, para esto se debe tomar en cuenta la clasificación de los datos de acuerdo al tipo de formato (textos, números, imágenes); al proceso de obtención (experimentales y observacionales) y al tipo de recolección de datos que se han utilizado. (Ver tabla 5. Clasificación de datos de investigación según la Reserach Information Network RIN (p.44).

3.3.3. Preservación de los datos de investigación

En esta etapa del modelo se debe realizar un plan de mantenimiento y fiabilidad de los medios de almacenamiento para todo el ciclo de vida de los datos. Este plan incluye la correcta administración de todo el ciclo de vida de los datos tomando en cuenta cuándo y cuáles van a ser desechados o trasladados a otro archivo, depósito, centro de datos o destruido de acuerdo a políticas de depuración a corto, mediano y largo plazo tomando como base el interés académico y administrativo de los mismos, así como también la verificación de la integridad de los datos y copias de seguridad a fin que estos sean catalogados y diseminados. La propuesta de este modelo sugiere que el gestor de la información realice el acompañamiento en todo momento al investigador durante todo el ciclo de vida de los datos como curador de datos (ver anexo I: glosario de términos), es decir realizando todas las actividades de la curación de datos (ver anexo I: glosario de términos), quienes tienen como función principal la capacitación e inducción en actividades de selección, registro, clasificación y almacenamiento de los datos con el fin de dar valor agregado a los mismos para que estos sean reutilizados. Estas actividades alcanzan todos los procesos que van desde la creación de los datos hasta que se determina que ya no son útiles, es decir que ya no cuentan con interés académico y/o científico. Pérez (2012). (p.45).

3.3.4. Almacenamiento de los datos

Esta etapa del ciclo de vida de los datos está referido principalmente al tipo de plataforma informática que se va a utilizar y configurar con el fin que todos los datos que se encuentran dispersos en diferentes dispositivos electrónicos de los investigadores puedan ser accedidos desde un solo punto de acceso para que puedan ser compartidos, reutilizados y preservados. Por tal motivo la estructura plantea que la plataforma de gestión cuente con un sistema de almacenamiento y análisis fiable y escalable, que tenga un subdominio propio, contenga el protocolo OAI-PMH (Ver anexo I: glosario de datos), para permitir la interoperabilidad con otras plataformas utilizadas para la gestión de los datos y que el diseño de la interfaz sea amigable a fin de identificar rápidamente las áreas de interés para su recuperación. Así mismo, se debe establecer un proceso de almacenamiento determinado en el plan de gestión de

datos que permita un proceso sencillo para que los materiales sean almacenados y conservados por los investigadores a fin de poder justificar los resultados de la investigación y permitir su reutilización. Cabe mencionar que en esta propuesta no se incluirá la implementación del repositorio de datos, debido a que esta configuración e implementación será realizada, por las áreas que determinen ambas universidades.

3.3.5. Acceso y uso de los datos

En esta etapa para poder acceder a los datos de investigación producidos por los investigadores y permitir la gestión de los datos, se considera importante que los datos que se publiquen sean originales, es decir que deben mantener una estructura que permita su uso automatizado como: lenguaje de etiquetado extensible: xml; infraestructura para descripción de recursos: rdf y valores separados por coma: csv. Así mismo, para que los investigadores conozcan lo que se ha publicado, se debe crear un catálogo en línea de datos que permitan contrastar la calidad de la información aportada. Los datos deben ser fiables, encontrarse estructurados y documentados y ser enriquecidos con herramientas de la web semántica para permitir compartir los datos de manera más segura. Así mismo, debe contar con estándares de esquemas de metadatos y softwares internacionales especializados en datos abiertos y que cuenten con códigos abiertos para que los datos puedan ser diseminados, compartidos y reutilizados por los investigadores (catalogación con metadatos apropiados y acordada por la comunidad que los produce).

3.3.6. Reutilización

En esta etapa del ciclo de vida de los datos, la estructura propuesta utiliza estándares internacionales de metadatos y softwares interoperables, que permiten comprender el contenido, formatos y contexto de los datos para que puedan ser utilizados y reutilizados por los investigadores como, en el caso de metadatos: el esquema de metadatos DataCite, Starr et al. 2015. (p.5) y la Guía de OpenAire. (Ver cuadro comparativo de esquemas de metadatos más utilizados para la gestión de datos de investigación: tabla 12). Para la descripción de la metadata se utilizará el modelo

de metadatos Dublin Core para describir los recursos, que cuenta con estándares internacionales de interoperabilidad y permitir la interoperabilidad de los mismos.

Tabla 12 Comparación de esquemas de metadatos más utilizados para la gestión de datos de investigación

Detalle	DataCite	Open Aire
Características	Acceso fácil a los datos y compatible.	Publicación de metadatos fácil y compatible. Adopta algunas normas de DataCite
Contribuciones citables	✓	Utiliza los señalados en el esquema de DataCite
Comprobación de los resultados y reutilización	✓	✓
Utiliza identificadores de objetos digitales	DOI	Utiliza DOI del esquema de Data Cite
Niveles de descripción de metadatos: Obligatorios, Recomendados y Opcionales	✓	✓
Utiliza protocolos de interoperabilidad	✓	✓
Utiliza metadatos de relación (dc.relation) que permite la reutilización de los datos	✓	✓

Fuente: elaboración propia

Si bien los esquemas de metadatos descritos en la tabla de arriba poseen características similares, para el modelo a utilizar en el presente estudio se ha seleccionado el uso de OpenAire como el esquema más adecuado debido a que utiliza Dublin Core para la descripción de los recursos. Este esquema es el propuesto por LA Referencia, Red de Repositorios de Latinoamérica quien es la encargada de dar las normas y uso de estándares para los repositorios institucionales a nivel regional y el CONCYTEC en el Perú. En relación a los formatos a utilizar, se sugiere utilizar las hojas de cálculo como XLS, CSV o XML y PDF para la entrega de diccionario de datos. (Ver tabla 13). A nivel nacional el CONCYTEC exige la entrega de un diccionario de datos junto con los datos que se van a publicar y en relación a las licencias para manejar códigos abiertos se recomienda las que son compatibles a datos abiertos (Ver tabla 14).

Tabla 13 Formatos recomendados para la reutilización de datos de investigación

Ejemplos de formatos recomendados para la reutilización de datos de investigación	
Formato	Función
XLS/ODS	Hojas de cálculo que muestran la información en celdas organizadas en filas y columnas
CSV	Representa datos en forma de tabla. Las columnas se separa por comas o tabulaciones y las filas por salto de línea.
XLM	Lenguaje general de etiquetado que define un conjunto de reglas para la codificación de documentos a través de modelos específicos legibles y procesables.
PDF	Formato utilizado para la entrega del diccionario de datos junto con los datos abiertos.

Fuente: España. Ministerio de Industria, e. y T.E (2015)

Tabla 14 Licencias compatibles con los principios de datos abiertos

Licencias compatibles con los principios de datos abiertos		
Nombre	Descripción	Tipo de licencia
Dominio público	Permiten compartir los datos sin ningún tipo de restricción.	<ul style="list-style-type: none"> • Creative Commons Zero (CCZero) • Public Domain Dedication and Licence (PDDL 1.0)
Atribuir la fuente original de los datos	Obligan a atribuir la fuente original de los datos	<ul style="list-style-type: none"> • Creative Commons-Reconocimiento (CC BY 4.0). • Attribution Licence 1.0 (ODC-BY 1.0)
Atribución de la fuente original y compartir los resultados	Requieren atribución de la fuente original y compartir los resultados bajo las mismas condiciones	<ul style="list-style-type: none"> • Creative Commons. Reconocimiento Compartir Igual (CC BY-SA 4.0) • Open Data Commons Open Database License ODbL 1.0) • Open Data Licence Atribucion

Fuente: Elaboración propia

En relación a los softwares internacionales a utilizar, se recomiendan el uso de CKAN o DKAN. Ambos softwares son de aplicación web de código abierto y han sido concebidos para la publicación de datos abiertos, exclusivamente. En esta propuesta se ha seleccionado el software DKAN por sus características que se adecúan de mejor manera al modelo propuesto de gestión de los datos de investigación y se adapta al modelo responsive para dispositivos móviles. Respecto a la gestión de los datos

abiertos hemos tomado en cuenta las políticas y lineamientos del gobierno peruano respecto al uso de los software para este fin. Actualmente el software que recomienda el CONCYTEC es el DKAN. Este software es utilizado actualmente para el portal de datos abiertos del Ministerio de Educación y para el portal de datos abiertos del Ministerio de Salud (ver tabla 15).

Tabla 15 Comparación de softwares de código abierto para la publicación de datos abiertos

Softwares de código abierto		
Características	CKAN	DKAN
Detalle	Permite publicar, almacenar y distribuir colecciones de datos con muy pocos recursos.	Permite la publicación, almacenamiento y distribución de colecciones de datos.
Control de contenidos	Sistema propio tablas.	Utiliza Drupal
Beneficios para los usuarios	Pueden explorar los datos y previsualizarlos a través de mapas, gráficas y tablas.	Ofrece navegación muy amigable con diferentes facetas de búsqueda
Estándares internacionales de datos abiertos	✓	✓
Visualización e interfaz	Muy amigable	Amigable

Fuente. Elaboración propia

3.3.7. Transformación de los datos

Creación de nuevos datos a partir de los originales a un formato diferente, a fin de generar nuevos resultados que dan inicio a un nuevo ciclo de vida de los datos y reducir la duplicación de esfuerzos en investigación. En esta etapa se deben extraer los datos almacenados para revisarlos, analizarlos y definir aquellos que sean utilizados para su explotación y utilización como: formato, nivel de acceso, diseminación, entre otros. Hay que realizar también modificaciones y revisiones del entorno e infraestructura para implementar los requerimientos y dar inicio a un nuevo ciclo de vida de los datos.

3.3.8. Difusión de los datos de investigación

Consideramos que la etapa de la difusión en el ciclo de vida de los datos es crítica debido al desconocimiento que tienen los investigadores sobre los beneficios de

compartir y reutilizar los datos de investigación, con la finalidad de aumentar el número de investigaciones, mejorar la calidad, economizar recursos y generar nuevos conocimientos. Las actividades de difusión, requiere de algunas estrategias básicas tales como:

- a. Utilizando la base de datos, se determina el número de potenciales usuarios.
- b. Elaborar mensajes con una comunicación de acuerdo al tipo de usuario: alumno, docente o investigador, y de acuerdo al recurso electrónico utilizado: web, redes sociales, o correo directo o cuenta del usuario UNMSM y/o UPC;
- c. El programa de difusión requiere de: un programa de inducción, a cada docente investigador interesado en compartir sus datos de investigación; un programa de capacitación colectivo; y un programa de publicidad por los sitios web o blogs más usados por los docentes investigadores, por correo electrónico, así como, por las redes sociales principalmente facebook, twitter, linkedin y youtube;
- d. Atender las solicitudes de inducción, acompañamiento permanente y absolver dudas y preguntas.

Para asegurar que las estrategias básicas de difusión despierten el interés, en los docentes investigadores, en compartir y reutilizar los datos de investigación se requiere puntualizar y repetir sistemáticamente los beneficios del acceso abierto a la información científica. Como parte de este plan se requiere de realizar actividades de: medición sistemática de las respuestas a los mensajes; seguimiento de los procesos de inducción y capacitación; comparación de estándares de las redes sociales; y eliminación de errores en cada uno de los programas.

Así mismo, para la reutilización de los datos, si bien sabemos que los investigadores desean compartir sus datos, no los comparten en su totalidad, debido a que no tienen la suficiente garantía para que sus datos se utilicen de forma responsable. Por tal motivo, el modelo propuesto coloca en todos sus registros la licencia Creative Commons para que de esa manera puedan ser fácilmente citados y reutilizados previa coordinación con los investigadores, y con las autoridades de la institución. Sin embargo, en el Perú el CONCYTEC recomienda el uso de la licencia Open Data Commons Open Database License ODbL 1.0)

CAPÍTULO IV: METODOLOGÍA

En este capítulo describimos la metodología empleada para recoger y obtener la información necesaria que ha servido para conseguir los resultados del presente estudio. Se explican: el tipo de investigación; el diseño de la investigación; las variables de estudio; operacionalización de las variables; la selección de la población; la técnica de recolección de datos; y el instrumento que conduce a obtener los resultados para su validación y análisis. El propósito de utilizar esta metodología es conseguir la información necesaria que nos permita diagnosticar cual es la situación actual de la gestión de los datos de investigación y elaborar una propuesta.

4.1. Tipo y Diseño de investigación

El tipo y diseño de la investigación es descriptiva, usando métodos cualitativos en el estudio de la prevalencia del comportamiento de los datos, no experimental y prospectivo, lo que contribuyó a la selección de los investigadores de la Universidad Nacional Mayor de San Marcos (UNMSM) y Universidad Peruana de Ciencias Aplicadas (UPC). Se determinó utilizar la técnica de encuesta y como instrumento un cuestionario adecuado.

4.2. Matriz de operacionalización de variables

Para elaborar el cuestionario hemos desarrollado la siguiente matriz de operacionalización, (Ver Tabla 16). La matriz permitirá medir las variables de la presente investigación a través de indicadores.

Tabla 16 Matriz de operacionalización de variables

Variables	Indicadores	Preguntas
Uso de los datos de investigación	<p>Conservación de datos de investigación</p> <p>Duración de la conservación de los datos de investigación.</p> <p>Disponibilidad de datos de investigación.</p> <p>Acceso a entidades.</p> <p>Medios de acceso a los datos de investigación por otros investigadores.</p> <p>Datos generados utilizados por otros investigadores-</p> <p>Reutilización de datos de investigación.</p>	<p>1. ¿Qué hace usted con sus datos de investigación?</p> <p>2. ¿Cuánto tiempo conserva sus datos de investigación desde que los recoge?</p> <p>3. ¿Cuánto de sus datos de investigación los coloca disponibles a los investigadores?</p> <p>4. ¿A quién concede usted el acceso a sus datos de investigación?</p> <p>5. ¿A través de que medio pueden acceder otros investigadores a los datos de su investigación?</p> <p>6. ¿Utiliza los datos generados por otros investigadores o instituciones cómo OMS u OPS ?</p> <p>7. ¿Reutiliza usted los datos de otros investigadores?</p>
Organización de los datos	<p>Software estándar</p> <p>Uso de formatos de almacenamiento</p> <p>Técnicas de recolección de datos.</p> <p>Tipo de datos de investigación</p>	<p>1. ¿Cómo registra Usted los datos de su investigación?</p> <p>2. ¿Qué tipo de formato usa cuando almacena datos de investigación?</p> <p>3. ¿Qué técnicas de recolección de datos utiliza?</p> <p>4. ¿Qué tipo de datos de investigación, utiliza Ud., en sus investigaciones?</p>

Almacenamiento de datos de investigación	<p>Condiciones de envío de datos de investigación</p> <p>Espacio de almacenamiento utilizado por un investigador.</p> <p>Tipo de almacenamiento más utilizado por los investigadores.</p> <p>Razones de disponibilidad y preservación</p> <p>Opciones de apoyo respecto a la implementación del repositorio de datos de investigación.</p> <p>Responsable de la gestión del archivo de datos de investigación.</p>	<ol style="list-style-type: none"> 1. ¿Qué condiciones requeriría para enviar sus datos de investigación a un repositorio de datos? 2. Estime el volumen total de sus datos de investigación basado en el espacio de almacenamiento que necesite. 3. ¿Qué tipo de almacenamiento utiliza para depositar sus datos de investigación? 4. En su opinión, cuáles son las razones más importantes para que los datos de investigación estén disponibles y preservados? 5. ¿Si su organización implementara un repositorio de datos de investigación, para la gestión de los mismos, qué opciones de apoyo esperaría Ud. de ella? 6. ¿Quién le gustaría que sea el responsable del archivo de sus datos de investigación en un repositorio de datos?
---	--	--

Fuente: Elaboración propia

4.3. Unidad de análisis

Se identificó a cada uno de los docentes investigadores del área de Ciencias de la Salud de las universidades San Marcos y UPC, a quienes se le entregó el cuestionario con 19 preguntas, para precisar la recolección de datos. En San Marcos los docentes pertenecen a la Facultad de Medicina; de Farmacia y Bioquímica; Odontología; y Psicología que tienen publicaciones en Web of Sciences (WOS), y/o SCOPUS. Y en la UPC los docentes investigadores son de Facultad de Ciencias de la Salud; y

Psicología, que también tienen publicaciones en Web of Sciences (WOS), y/o SCOPUS.

4.4. Población

La encuesta se hizo vía online a 80 docentes investigadores en Ciencias de la Salud: 40 docentes investigadores de la Universidad Nacional Mayor de San Marcos (UNMSM), y 40 docentes investigadores de la Universidad Peruana de Ciencias Aplicadas (UPC). Al cierre de la encuesta (25 de noviembre) se recibió respuesta de 9 docentes investigadores de la USMSM y 16 docentes investigadores de la UPC.

4.5. Muestra

La muestra para la encuesta fue de 25 docentes investigadores, que estuvo compuesta por: 9 docentes investigadores de la Universidad Nacional Mayor de San Marcos (UNMSM), y 16 docentes investigadores de la Universidad Peruana de Ciencias Aplicadas (UPC).

4.6. Técnica de recolección de datos

Con el propósito de conocer el proceso de gestión de datos de investigación, principalmente sobre el uso, organización y almacenamiento, en el área de Ciencias de la Salud de dos universidades peruanas se aplicó la técnica de la encuesta para la recolección de datos.

Se utilizó como instrumento un cuestionario de 19 preguntas cerradas en el formulario Google Form, por su eficacia y confiabilidad respecto a los resultados. El cuestionario fue dividido en cuatro bloques: un bloque sobre información general de los investigadores; y tres bloques sobre las variables de estudio: uso; organización; y almacenamiento de los datos de investigación.

El instrumento se elaboró de acuerdo al perfil de los docentes investigadores de las áreas de Ciencias de la Salud de las universidades UNMSM y UPC. Fue remitido vía online bajo las siguientes direcciones: UNMSM <https://goo.gl/jQFJGY> y UPC

<https://goo.gl/kTvfup>. El instrumento se puso a disposición de los investigadores durante 10 días consecutivos

Se debe aclarar que el cuestionario es un instrumento validado y aplicado en diversas encuestas realizadas por instituciones europeas y norteamericanas con gran experiencia en gestión de datos de investigación. Estas encuestas son: Carol T., et.al. (2015). *Changes in Data Sharing and Data Reuse Practices and Perceptions Among Scientists Worldwide*; Federer; L. M., et.al (2015). *Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff*; J Thorp. (2015); *Questionnaire National Research Data Survey*; Luzi, D., Ruggieri, et.al (2013). *Data sharing in environmental sciences: A survey of CNR researchers*.

4.7. Elaboración del cuestionario

Se elaboró un cuestionario de 19 preguntas cerradas, agrupándose en 4 bloques con los siguientes contenidos específicos:

Bloque I: información general, preguntas del 1 al 3, este bloque se refiere al lugar donde el investigador realiza su investigación, la línea o líneas de su investigación, y de donde provienen los fondos de su financiamiento;

Bloque II: uso de los datos de investigación, preguntas del 4 al 9, este bloque se refiere al uso que el investigador da a sus datos, cómo, dónde y cuánto tiempo conserva sus datos de investigación, si los comparte, y cómo y qué medios utiliza para acceder a los datos de investigación de otros investigadores;

Bloque III: organización de datos de investigación, preguntas del 10 al 13, este bloque se refiere a aspectos organizativos y administrativos sobre registros, formatos, técnicas de recolección y el tipo de dato de investigación que utiliza; y el

Bloque IV: almacenamiento de datos de investigación, preguntas del 14 al 19, este bloque se refiere a las condiciones, volumen, tipo de almacenamiento, disponibilidad y como almacenar y preservar sus datos de investigación.

El cuestionario, ver Formato 1, es teóricamente justificado bajo los siguientes criterios:

- Contenido: las preguntas elegidas para el cuestionario, son adecuadas, fiables y han permitido medir las variables (uso, organización y almacenamiento de datos de investigación) que provienen de la matriz de operacionalización de las variables
- Utilidad: el cuestionario a través del formulario Google Form es útil para este tipo de encuestas por su reconocida eficacia y confiabilidad respecto a los resultados, esto debido a que se sostiene en una plataforma informática adecuada para esclarecer la situación actual sobre el uso, organización y almacenamiento de los datos de investigación.
- Flexibilidad: la elaboración del cuestionario se realizó tomando en cuenta el perfil del grupo objetivo (docentes investigadores de las áreas de Ciencias de la Salud de las universidades UNMSM y UPC), adaptándose las preguntas al contexto universitario.
- Personalizado: el envío del cuestionario fue personalizado vía online bajo las siguientes direcciones: UNMSM <https://goo.gl/jQFJGY> y UPC <https://goo.gl/kTvfUP>, lo que garantizó la recepción del cuestionario por el docente investigador; y
- Disponibilidad: el cuestionario estuvo a disposición de los investigadores por 10 días consecutivos y se basa en la aceptación voluntaria a participar en la encuesta de los 25 docentes investigadores que cuentan con la experiencia académica y científica para dar respuestas con criterio personal.

Formato 1 Encuesta

I. INFORMACIÓN DEL INVESTIGADOR(A)

1. Seleccione la institución en la que realiza trabajos de investigación. *

Por favor, escoja todas las opciones que apliquen.

- ☐ Universidad Nacional Mayor de San Marcos
- ☐ Universidad Peruana de Ciencias Aplicadas
- ☐ Otro:

2. ¿Cuáles son las líneas de investigación a las que usted se dedica?*

Por favor, escoja todas las opciones que apliquen.

- ☐ Biomateriales
- ☐ Biología sintética y Bionanotecnologías
- ☐ Bioquímica, Biología Molecular y Celular
- ☐ Educación en Ciencias de la Salud
- ☐ Epidemiología Clínica y Salud Basada en Evidencias
- ☐ Nutrición, Deporte y Actividad Física
- ☐ Salud Pública y Gestión en Salud
- ☐ Personalidad y Desarrollo
- ☐ Psicología deportiva
- ☐ Psicología social y de las organizaciones
- ☐ Psicología clínica
- ☐ Otro:

3. ¿Cuál de las siguientes opciones describe mejor la principal agencia de fondos para su investigación?*

Por favor, escoja todas las opciones que apliquen.

- ☐ Gobierno nacional
- ☐ Entidad privada
- ☐ Mi institución

II. USO DE DATOS DE INVESTIGACIÓN

4. ¿Qué hace usted con sus datos de investigación?*

Por favor, escoja una opción

- ☐ Los conserva
- ☐ Los desecha

En caso su respuesta sea "Los conserva", ¿Cuánto tiempo los conservaría?

Por favor, escoja una opción

- ☐ Más de 1 año
- ☐ Menos de 1 año

5. ¿Cuánto de sus datos de investigación los coloca disponibles a los demás?*

Por favor, escoja una opción

- ☐ Todos
- ☐ Algunos
- ☐ Ninguno

6. ¿A quién concede usted el acceso a sus datos de investigación? *

Por favor, escoja todas las opciones que apliquen.

- ☐ Al público
- ☐ A la comunidad científica
- ☐ A los investigadores de mi misma disciplina
- ☐ A todos los miembros de mi institución
- ☐ A las personas interesadas en solicitar mis datos

7. ¿A través de que medio pueden acceder otros investigadores a los datos de su investigación?*

Por favor, escoja todas las opciones que apliquen.

- ☐ Repositorio de datos
- ☐ Como material complementario vinculado a sus publicaciones
- ☐ Web personal o institucional
- ☐ Dispositivos electrónicos (USB, CD-Rom, etc) y/o correo electrónico

8. ¿Utiliza los datos generados por otros investigadores o instituciones como OMS u OPS? *

Por favor, escoja una opción

- ☐ Si
- ☐ No

9. ¿Reutiliza usted los datos de otros investigadores?*

De acuerdo a la definición recogida de la encuesta Questionnaire National Research Data Survey (2013), "Reutilizar significa que los datos pueden ser utilizados y tratados posteriormente teniendo en cuenta las reglas predefinidas de lo que está permitido hacer con los datos utilizando por ejemplo licencias de acceso abierto". Por favor, escoja una opción

- ☐ Si
- ☐ No

En caso su respuesta sea afirmativa, seleccione el tipo de licencia que aplicaría.

Por favor, escoja una opción

- ☐ Creative Commons
- ☐ Open Data License Attribution
- ☐ Acuerdos de cooperación
- ☐ Ninguno
- ☐ No sé
- ☐ Otro:

III. ORGANIZACIÓN DE DATOS DE INVESTIGACIÓN

10. ¿Cómo registra usted los datos de su investigación? *

Por favor, escoja todas las opciones que apliquen.

- ☐ Utilizando un software adecuado
- ☐ Utilizando alguna nomenclatura
- ☐ Utilizando estándares internacionales
- ☐ No utilizo
- ☐ Otro:

11. ¿Qué tipo de formato usa cuando almacena datos de investigación?*

Por favor, escoja todas las opciones que apliquen.

- ☐ Documentos de texto (DOC, ODF, TXT, etc.)
- ☐ Documentos Digitales (PDF, LATEX, etc.)
- ☐ Texto estructurado (HTML, JSON, TEX, XML, etc.)
- ☐ Hojas de cálculo (XLS, ODS, CSV, SAS, etc.)
- ☐ Bases de datos (MS Access, MySQL, Oracle, NON-SQL, etc.)
- ☐ Gráficos / Imágenes (JPEG, SVG, PNG, GIF, TIFF, etc.)
- ☐ Audio (MP3, WAV, AIFF, OGG, etc.)
- ☐ Vídeos / películas (MPEG, AVI, WMV, MP4, etc.)
- ☐ Software de código de las aplicaciones de código (CSS, JavaScript, Java, etc.)
- ☐ Datos de configuración (INI, CONF etc.)
- ☐ Aplicaciones de software
- ☐ Otro:

12. ¿Qué técnicas de recolección de datos utiliza?*

Por favor, escoja todas las opciones que apliquen.

- ☐ Experimental (implica un cierto grado de manipulación)
- ☐ Entrevista
- ☐ Observacional (no incluida la manipulación)
- ☐ Encuesta
- ☐ Otro:

13. ¿Qué tipo de datos de investigación utiliza usted en sus investigaciones. *

Por favor, escoja todas las opciones que apliquen.

- ☐ Datos numéricos
- ☐ Resultados de la medida de instrumentos
- ☐ Datos de encuestas
- ☐ Imágenes digitales (fotografía, diagramas)
- ☐ Audios digitales (archivos sonoros, entrevistas)
- ☐ Documentación diversa
- ☐ Otro:

IV. ALMACENAMIENTO DE DATOS DE INVESTIGACIÓN**14. ¿Qué condiciones requeriría para enviar sus datos de investigación a un repositorio de datos? ***

Por favor, escoja todas las opciones que apliquen.

- ☐ Confidencialidad y seguridad
- ☐ Capacidad para actualizar los datos
- ☐ Capacidad para eliminar los datos
- ☐ Saber quién está utilizando los datos, cuándo y para qué fin
- ☐ Contactar con alguien que requiera utilizar mis datos
- ☐ Recibir un reconocimiento formal
- ☐ Preservación de datos a largo plazo
- ☐ Procedimientos sencillos para el registro de los datos
- ☐ Recibir fondos adicionales
- ☐ Recibir la misma evaluación recibida para su publicación en una revista
- ☐ Otro:

15. Estime el volumen total de sus datos de investigación basado en el espacio de almacenamiento que necesite (el promedio estimado por investigación).*

Por favor, escoja una opción

- ☐ Más de un Tera
- ☐ Menos de un Tera

16. ¿Qué tipo de almacenamiento actualmente utiliza para depositar sus datos de investigación? *

Por favor, escoja todas las opciones que apliquen.

- ☐ En un centro de datos externo
- ☐ En un servicio en la nube
- ☐ En el repositorio institucional de la universidad
- ☐ En la computadora de mi trabajo
- ☐ En mi computadora personal
- ☐ En un disco duro externo (también unidad USB)
- ☐ En CD / DVD
- ☐ En otro repositorio de datos
- ☐ En el repositorio de una disciplina determinada (Por ejemplo: NEON o LTER)
- ☐ En servicios de almacenamientos de datos como GITHUB
- ☐ Otro:

17. En su opinión ¿Cuáles son las razones más importantes para que los datos de investigación estén disponibles y preservados? *

Por favor, escoja todas las opciones que apliquen.

- ☐ Mejora la transparencia de los resultados de la investigación.
- ☐ Es un medio para validar los resultados obtenidos
- ☐ Disponibilidad de datos, fomenta el progreso de la ciencia (nueva investigación se basa en el conocimiento preexistente)
- ☐ Los resultados existentes se pueden volver a examinar.
- ☐ Promueve la colaboración entre campos de diferentes disciplinas.
- ☐ Tiene un valor económico potencial
- ☐ Los datos de investigación son únicos
- ☐ La disponibilidad de datos reduce la duplicación de esfuerzos de investigación

18. ¿Si su organización implementara un repositorio de datos de investigación para la gestión de los mismos, qué opciones de apoyo esperaría usted de ella? *

Por favor, escoja todas las opciones que apliquen.

- ☐ Directrices o políticas sobre la gestión de datos de investigación
- ☐ Incluir un plan de gestión de datos de investigación.
- ☐ Cursos de capacitación o entrenamiento
- ☐ Gestor de la información especializado en el procedimiento de datos de investigación
- ☐ Apoyo en la elección de estándares, licencias de acceso abierto y metadatos.
- ☐ Otro:

19. ¿Quién le gustaría que sea el responsable del archivo de sus datos de investigación en un repositorio de datos? *

Por favor, escoja todas las opciones que apliquen.

- ☐ Yo mismo
- ☐ Yo mismo, pero acompañado de un experto en datos de investigación (data curator)
- ☐ Centro de TI
- ☐ Gestor de la Información
- ☐ Otro:

CAPÍTULO V: ANÁLISIS E INTERPRETACIÓN DE LOS RESULTADOS

El análisis y la interpretación de los resultados obtenidos estarán enfocados en el uso de los datos de investigación; la organización de los datos de investigación y el almacenamiento de los datos de investigación. Para este fin utilizaremos una tabla de frecuencia donde se muestran las frecuencias absolutas y las frecuencias relativas de las respuestas del cuestionario respondido por los docentes investigadores, donde:

- **Frecuencia absoluta:** se define como el número de veces que se repite la misma respuesta o dato observado, que cumple unas características determinadas y se denotará como “**fi**”; y
- **Frecuencia relativa:** es la proporción en términos de porcentaje (%) que representa “**fi**” en el total de respuestas recibidas o datos observados “**N**”, se denotará por “**hi**” y se calcula con la siguiente fórmula: $hi = (fi/N)$.

5.1. Análisis e interpretación de los resultados UNMSM

A continuación se realiza el análisis de los resultados de las 19 preguntas de la encuesta realizada en la UNMSM a 9 docentes investigadores en Ciencias de la Salud (CS). Tomándose en cuenta las frecuencias absolutas o el número de respuestas que se repiten y las frecuencias relativas o la participación de cada respuesta en el total de respuestas.

Bloque I: información general

Tabla 17 Distribución de frecuencias: instituciones donde los investigadores de CS de la UNMSM realizan sus trabajos de investigación

1. Seleccione la institución en la que realiza trabajos de investigación.

Respuesta	fi	hi
Universidad Nacional Mayor de San Marcos	9	75%
Universidad Peruana de Ciencias Aplicadas	1	8%
Instituto Nacional de Salud	1	8%
Instituto Nacional Cardiovascular	1	8%
Total	12	100%

Gráfico 18 Frecuencia relativa: instituciones donde los investigadores de CS de la UNMSM realizan sus trabajos de investigación



Se determinó que los investigadores utilizan más de una institución para realizar trabajos de investigación. Los resultados indican que los investigadores utilizan 4 instituciones. Se determinó que el 75% (9 trabajos de investigación) se realizan en la UNMSM y el 25% (3 trabajos de investigación) se hacen en instituciones privadas.

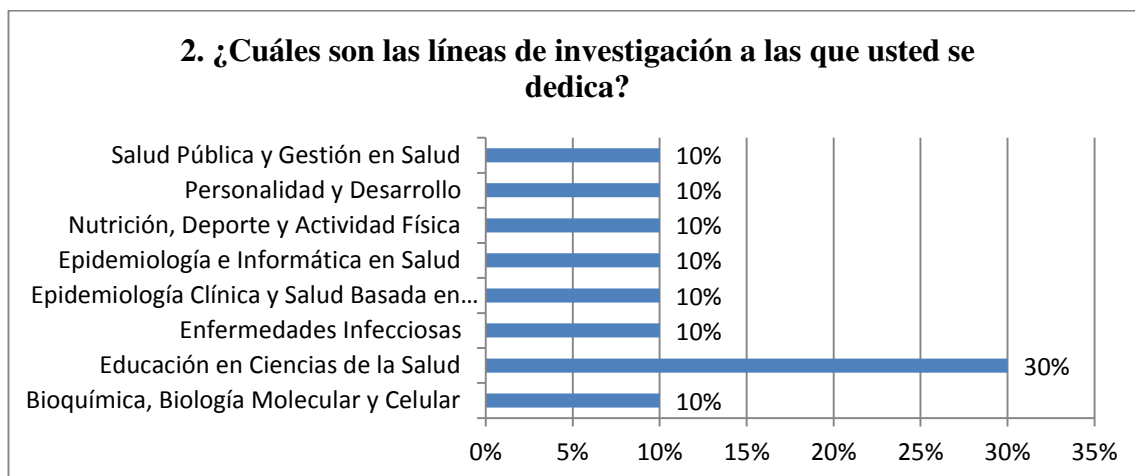
Este alto porcentaje de investigaciones que se realiza en la UNMSM se debe a las facilidades que los investigadores tienen por ser docentes en esa universidad. El que los investigadores realicen investigaciones en entidades privadas se debe principalmente a su trayectoria, su línea de investigación o porque son docentes investigadores en esas instituciones.

Tabla 18 Distribución de frecuencias: líneas de investigación a las que se dedican los investigadores de CS de la UNMSM

2. ¿Cuáles son las líneas de investigación a las que usted se dedica?

Respuesta	fi	hi
Bioquímica, Biología Molecular y Celular	1	10%
Educación en Ciencias de la Salud	3	30%
Enfermedades Infecciosas	1	10%
Epidemiología Clínica y Salud Basada en Evidencias	1	10%
Epidemiología e Informática en Salud	1	10%
Nutrición, Deporte y Actividad Física	1	10%
Personalidad y Desarrollo	1	10%
Salud Pública y Gestión en Salud	1	10%
Total	10	100%

Gráfico 19 Frecuencia relativa: líneas de investigación a las que se dedican los investigadores de CS de la UNMSM



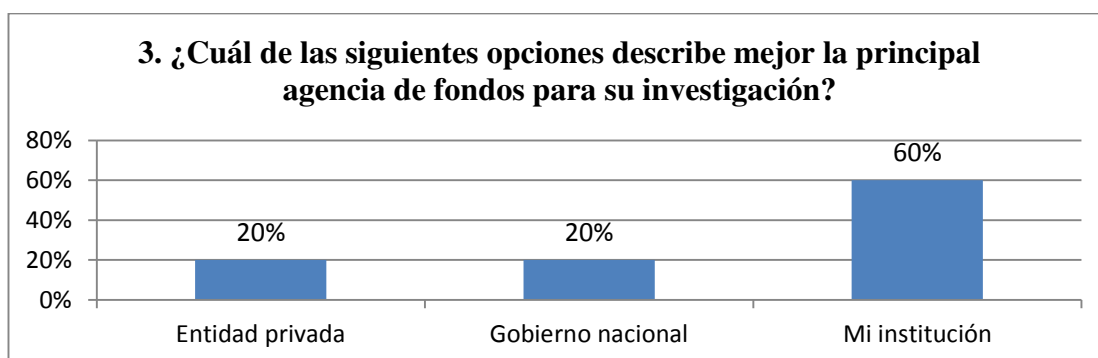
Se determinó que los investigadores tienen más de una línea de investigación. Asimismo, se determinó que son 8 las líneas de investigación en las que se desarrollan los docentes investigadores de Ciencias de la Salud en la UNMSM. El 30% (3 investigadores) realizan sus investigaciones en el campo de la Educación en Ciencias de la Salud.

Tabla 19 Distribución de frecuencias: agencias que contribuyen con fondos a la investigación en UNMSM

3. ¿Cuál de las siguientes opciones describe mejor la principal agencia de fondos para su investigación?

Respuesta	fi	hi
Entidad privada	2	20%
Gobierno nacional	2	20%
Mi institución	6	60%
Total	10	100%

Gráfico 20 Frecuencia relativa: las agencias que contribuyen con fondos a la investigación en UNMSM



Se determinó que los investigadores utilizan más de una agencia para proveerse de fondos para su investigación. Sobre el origen de los fondos de financiamiento de las investigaciones se determinó que el 60% (6 investigaciones) son financiadas por su institución. El 20% (2 investigaciones) son financiadas por entidades privadas y otro 20% (2 investigaciones) son financiadas por el gobierno nacional.

El financiamiento por la misma institución, se debe principalmente porque las universidades nacionales cuentan con convenios y fondos de origen internacional, que vienen dirigidos a líneas de investigación específicas. Respecto a los fondos que provienen de entidades privadas y del gobierno nacional, para financiar investigaciones, todavía son escasos.

Bloque II: uso de datos de investigación

Tabla 20 Distribución de frecuencias: uso que le da el investigador de CS de la UNMSM a sus datos investigación

4. ¿Qué hace usted con sus datos de investigación?

Respuesta	fi	hi
Los conserva	9	100%
Los desecha	0	0%
Total	9	100%

Gráfico 21 Frecuencia relativa: uso que le da el investigador de CS de la UNMSM a sus de datos de investigación

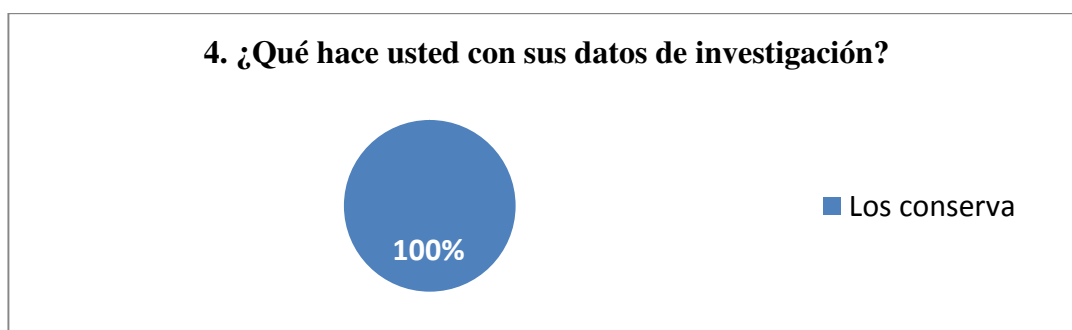
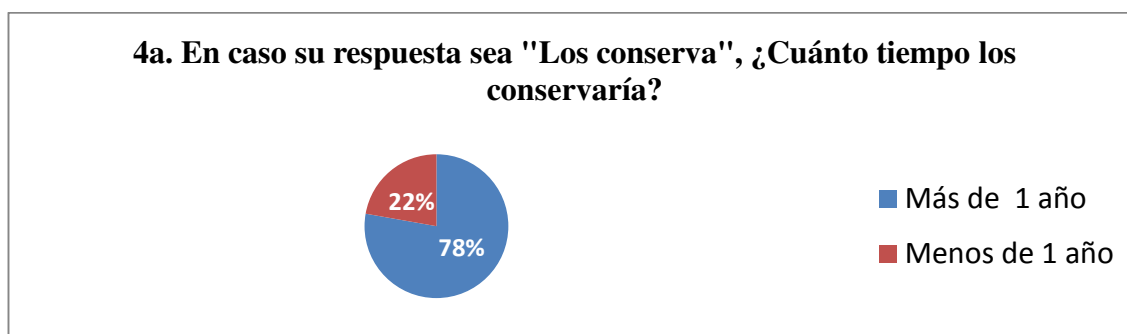


Tabla 21 Distribución de frecuencias: tiempo que conserva el investigador de CS de la UNMSM sus datos de investigación

4ª. En caso su respuesta sea “Los conserva”, ¿Cuánto tiempo los conservaría?

Respuesta	fi	hi
Más de 1 año	7	78%
Menos de 1 año	2	22%
Total	9	100%

Gráfico 22 Frecuencia relativa: tiempo que conserva el investigador de CS de la UNMSM sus datos de investigación



Se determinó que el 100% de los investigadores conserva sus datos de investigación. El 78% (7 investigadores) conserva sus datos de investigación por más de un año y sólo el 22% (2 investigadores) conserva sus datos de investigación por menos de un año.

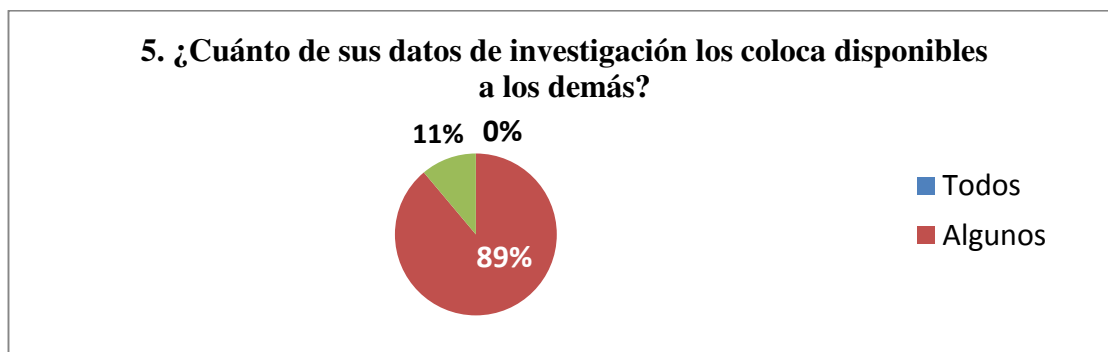
El conservar los datos de investigación se debe al sentido de previsión para el uso de los mismos en un futuro cercano o al sentido de protección de la propiedad intelectual.

Tabla 22 Distribución de frecuencias: cantidad de datos de investigación que el investigador de CS de la UNMSM coloca disponibles a los demás

5. ¿Cuánto de sus datos de investigación los coloca disponibles a los demás?

Respuesta	fi	hi
Todos	0	0%
Algunos	8	89%
Ninguno	1	11%
Total	9	100%

Gráfico 23 Frecuencia relativa: cantidad de datos de investigación que el investigador de la UNMSM pone a disponibilidad de los demás



Se determinó que ninguno de los investigadores coloca todos sus datos de investigación disponible para los demás. El 89% (8 investigadores) coloca sólo algunos de sus datos de investigación disponibles para los demás y el 11% (1 investigador) no coloca ningún dato disponible para los demás.

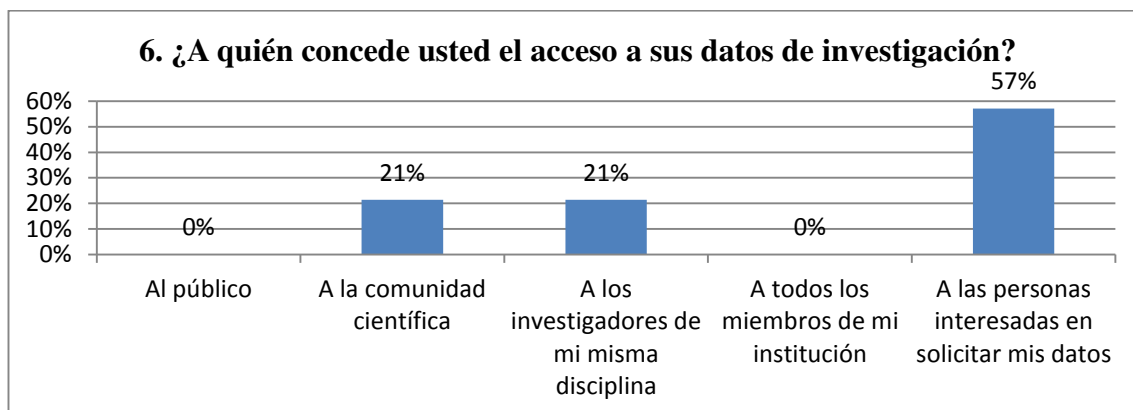
Esto se debe a que no prevalece entre los investigadores el hábito o la costumbre de colocar los datos de investigación para ser usado por otros investigadores, o que no encuentran una plataforma o un repositorio confiable o al temor de que se les dé, a los datos de investigación un mal uso.

Tabla 23 Distribución de frecuencias: usuarios que tienen acceso a los datos de investigación de los investigadores de CS de la UNMSM

6. ¿A quién concede usted el acceso a sus datos de investigación?

Respuesta	fi	hi
Al público	0	0%
A la comunidad científica	3	21%
A los investigadores de mi misma disciplina	3	21%
A todos los miembros de mi institución	0	0%
A las personas interesadas en solicitar mis datos	8	57%
Total	14	100%

Gráfico 24 Frecuencia relativa: usuarios que tiene acceso a los datos de investigación de los investigadores de CS de la UNMSM



Se determinó que los investigadores conceden a más de una entidad el acceso a sus datos de investigación. El 57% (8 investigadores) da acceso a sus datos de investigación a las personas que le solicitan esa información. Sólo el 21% (3 investigadores) concede el acceso a sus datos de investigación a la comunidad científica y otro 21% (3 investigadores) a los investigadores de la misma disciplina.

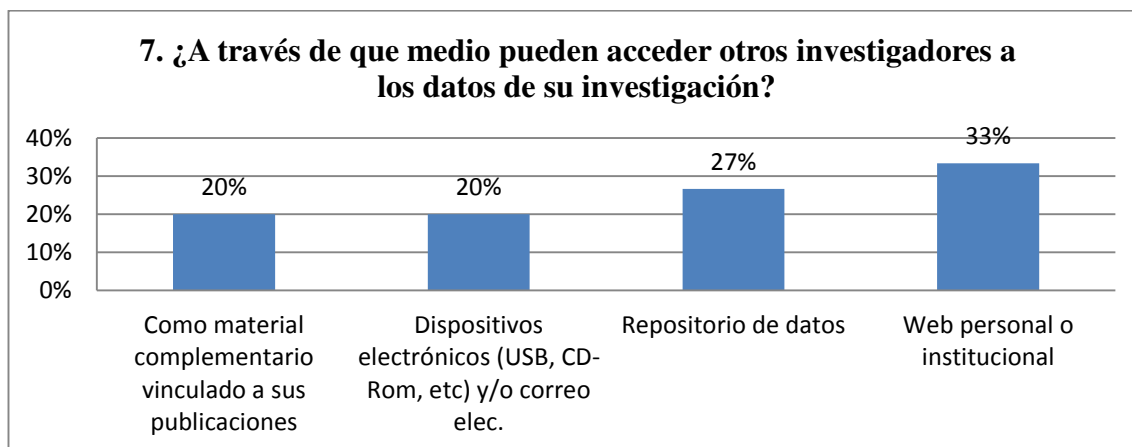
Cabe resaltar que ningún investigador da acceso, a sus datos de investigación, al público. Esto se debe a que no existe la confianza, de parte de los investigadores, para dar acceso a sus datos de investigación a cualquier persona, salvo solicitud expresa.

Tabla 24 Distribución de frecuencias: medios a través de los cuales se accede a los datos de investigación de los investigadores de CS de la UNMSM

7. ¿A través de que medio pueden acceder otros investigadores a los datos de su investigación?

Respuesta	fi	hi
Como material complementario vinculado a sus publicaciones	3	20%
Dispositivos electrónicos (USB, CD-Rom, etc) y/o correo elec.	3	20%
Repositorio de datos	4	27%
Web personal o institucional	5	33%
Total	15	100%

Gráfico 25 Frecuencia relativa: medios a través de los cuales se accede a los datos de investigación de los investigadores de CS de la UNMSM



Se determinó que los investigadores utilizan más de un medio para dar acceso a sus datos de investigación. Se determinó que el 33% (5 investigadores) dan acceso a sus datos de investigación a través de una web personal o institucional. El 27% (4 investigadores) lo hacen a través de un repositorio de datos. El 20% (3 investigadores) utilizan, como medio, dispositivos electrónicos, y otro 20% (3 investigadores) lo hacen como material complementario.

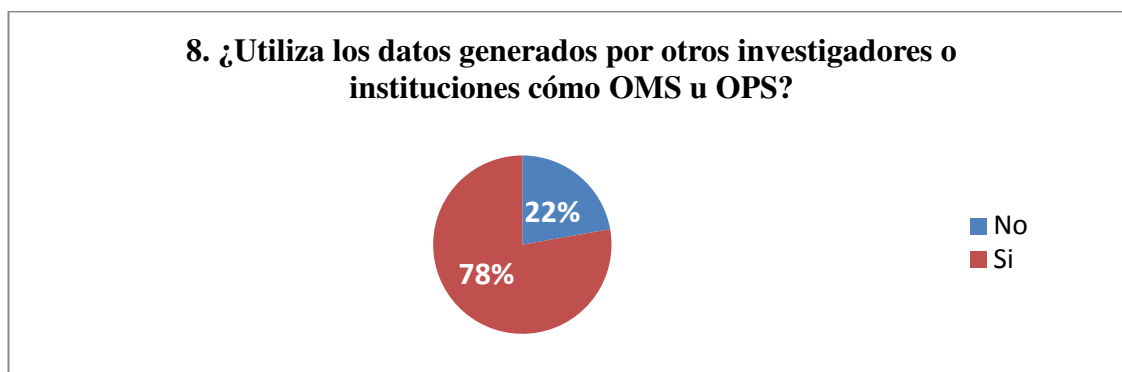
Es evidente que el uso de Internet ha facilitado el acceso a los datos e información ya que el 100% de accesos se dan por esa vía.

Tabla 25 Distribución de frecuencias: utilización de datos generados por otros investigadores o instituciones, realizados por los investigadores de CS de la UNMSM

8. ¿Utiliza los datos generados por otros investigadores o instituciones como OMS u OPS?

Respuesta	fi	hi
No	2	22%
Si	7	78%
Total	9	100 %

Gráfico 26 Frecuencia relativa: utilización de datos generados por otros investigadores o instituciones, realizados por los investigadores de CS de la UNMSM



Se determinó que el 78% (7 investigadores) utiliza los datos generados por otros investigadores o instituciones como la OMS u OPS. Y sólo el 22% (2 investigadores) no utilizan los datos generados por otros investigadores.

En el primer caso se debe a un categórico que los investigadores están en la misma línea de investigación de otros investigadores que publican y comparten sus datos de investigación. En el segundo caso se debe a que su línea de investigación es única o a las dificultades que tiene para acceder a los datos de investigación publicados.

Tabla 26 Distribución de frecuencias: reutilización de datos generados por otros investigadores o instituciones, realizados por los investigadores de CS de la UNMSM

9. ¿Reutiliza usted los datos de otros investigadores?

Respuesta	fi	hi
No	7	78%
Si	2	22%
Total	9	100%

Gráfico 27 Frecuencia relativa: reutilización de datos generados por otros investigadores o instituciones, realizados por los investigadores de CS de la UNMSM

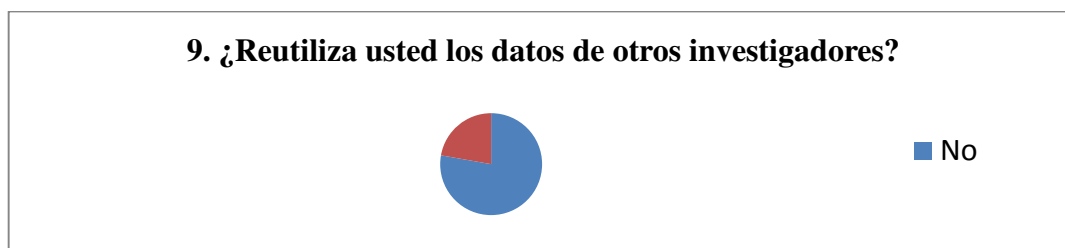
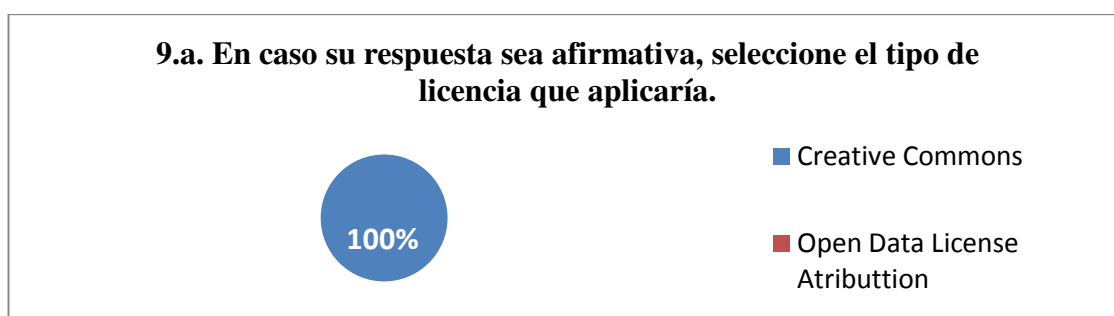


Tabla 27 Distribución de frecuencias: tipo de licencia que aplican los investigadores de CS de la UNMSM

9.a. En caso su respuesta sea afirmativa, seleccione el tipo de licencia que aplicaría.

Respuesta	fi	hi
Creative Commons	2	100%
Open Data License Attribution	0	0%
Acuerdos de cooperación	0	0%
Ninguno	0	0%
No sé	0	0%
Total	2	100%

Gráfico 28 Participación del tipo de licencia que aplican los investigadores de CS de la UNMSM



Se determinó que el 78% (7 investigadores) no reutiliza los datos de investigación de otros investigadores. Sólo el 22% (2 investigadores) reutiliza los datos de investigación de otros investigadores. Se determinó que el 100% (9 investigadores) utiliza la licencia Creative Commons.

La reutilización de datos de investigación se refiere a las facilidades que se dan para que un investigador pueda usar las veces que desee o que crea conveniente un

dato de investigación, que se encuentra en un repositorio o en una base de datos de acceso abierto, la finalidad de la reutilización es promover y facilitar la investigación y consecuentemente la innovación, racionalizando el uso de recursos escasos. En este caso, los investigadores no reutilizan los datos de investigación de otros investigadores, porque desconocen su existencia o no tienen los medios tecnológicos para hacerlo. Se determinó que el 100% (2 investigadores) que reutilizan datos de investigación recurren a la licencia Creative Commons.

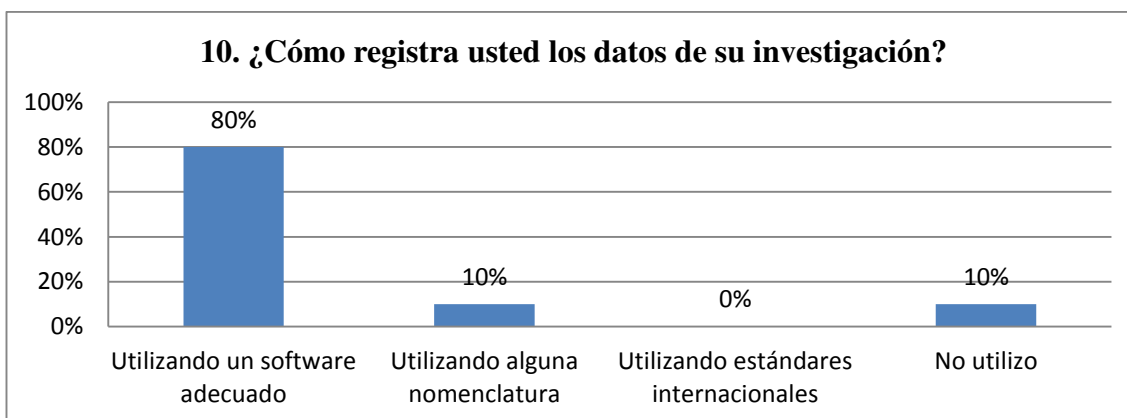
Bloque III: organización de datos de investigación

Tabla 28 Distribución de frecuencias: cómo registran sus datos de investigación los investigadores de CS de la UNMSM

10. ¿Cómo registra usted los datos de su investigación?

Respuesta	fi	hi
Utilizando un software adecuado	8	80%
Utilizando alguna nomenclatura	1	10%
Utilizando estándares internacionales	0	0%
No utilizo	1	10%
Total	10	100%

Gráfico 29 Frecuencia relativa: sobre como registran sus datos de investigación los investigadores de CS de la UNMSM



Se determinó que los investigadores utilizan más de un modo para registrar sus datos de investigación. Así tenemos, que el 80% (8 investigadores) registra sus datos utilizando un software adecuado, el 10% (1 investigador) registra sus datos utilizando

alguna nomenclatura y otro 10% (1 investigador) no utiliza ningún método para ese fin.

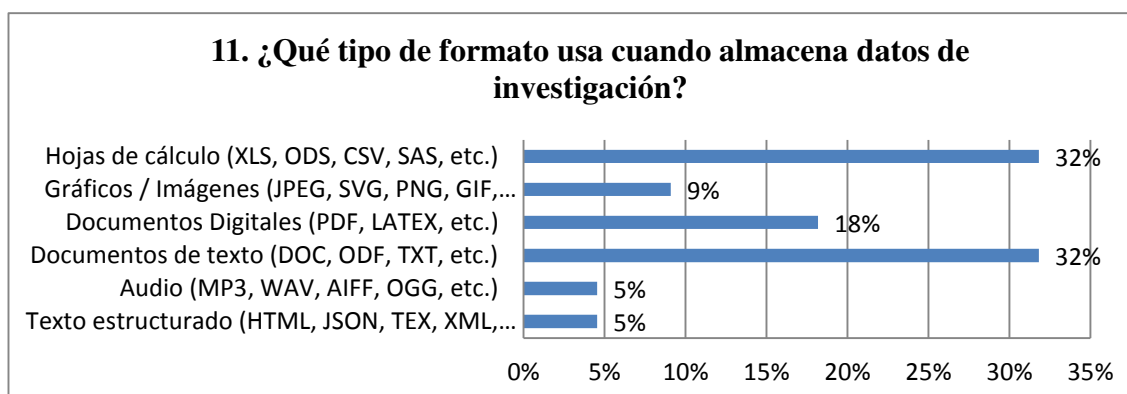
Se debe tomar en cuenta que la computadora es una herramienta básica para el trabajo de los investigadores y que por lo tanto debe tener un software actualizado para el registro temporal o permanente de sus datos de investigación.

Tabla 29 Distribución de frecuencias: tipo de formato que usa el investigador de CS de la UNMSM para almacenar sus datos de investigación

11. ¿Qué tipo de formato usa cuando almacena datos de investigación?

Respuesta	fi	hi
Texto estructurado (HTML, JSON, TEX, XML, etc.)	1	5%
Audio (MP3, WAV, AIFF, OGG, etc.)	1	5%
Documentos de texto (DOC, ODF, TXT, etc.)	7	32%
Documentos Digitales (PDF, LATEX, etc.)	4	18%
Gráficos / Imágenes (JPEG, SVG, PNG, GIF, TIFF, etc.)	2	9%
Hojas de cálculo (XLS, ODS, CSV, SAS, etc.)	7	32%
Total	22	100 %

Gráfico 30 Frecuencia relativa: tipo de formato que usa el investigador de CS de la UNMSM para almacenar sus datos de investigación



Se determinó que los investigadores utilizan más de un formato para almacenar sus datos de investigación. El 32% (7 investigadores) utiliza hojas de cálculo (XLS, ODS, CSV, SAS, etc.). Otro 32% (7 investigadores) utiliza documentos de texto

(DOC, ODF, TXT, etc.). El 18% (4 investigadores) utilizan documentos Digitales (PDF, LATEX, etc.), y el 9% (2 investigadores) utilizan formatos diversos como gráficos, e imágenes. El 5% (1 investigador) utiliza formatos de audio y otro 5% (1 investigador) utiliza textos estructurados.

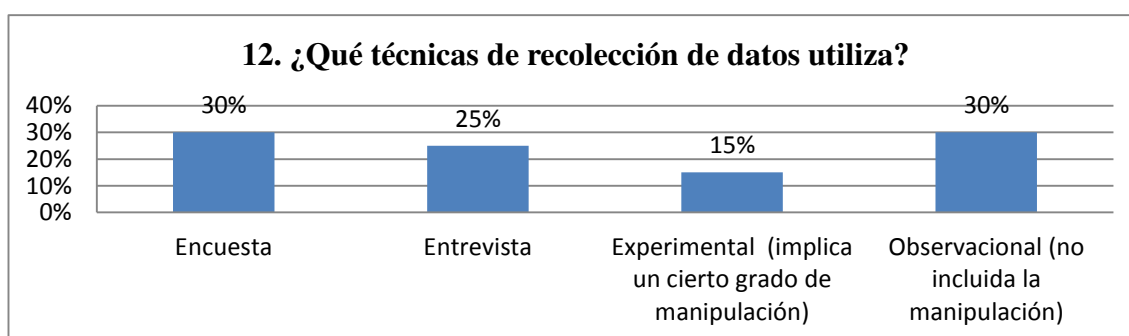
Como se observa, la mayoría de docentes investigadores utilizan formatos que son fáciles de usar, económicos y que están a su alcance a través del Office de Microsoft, siendo, algunos de ellos formatos adecuados para la reutilización de los datos, pero que no son utilizados como estándares en todas sus investigaciones para este fin,

Tabla 30 Distribución de frecuencias: técnicas de recolección de datos que utiliza el investigador de CS de la UNMSM

12. ¿Qué técnicas de recolección de datos utiliza?

Respuesta	fi	hi
Encuesta	6	30%
Entrevista	5	25%
Experimental (implica un cierto grado de manipulación)	3	15%
Observacional (no incluida la manipulación)	6	30%
Total	20	100%

Gráfico 31 Frecuencia relativa: técnicas de recolección de datos que utiliza el investigador de CS de la UNMSM



Se determinó que los investigadores utilizan más de una técnica de recolección de datos. Así mismo, se determinó que las técnicas de recolección de datos utilizada por la mayoría de investigadores son la observacional 30% (6 investigadores) y la encuesta 30% (6 investigadores). El 25% (5 investigadores) utiliza la entrevista y sólo el 15% (3 investigadores) utiliza la técnica experimental.

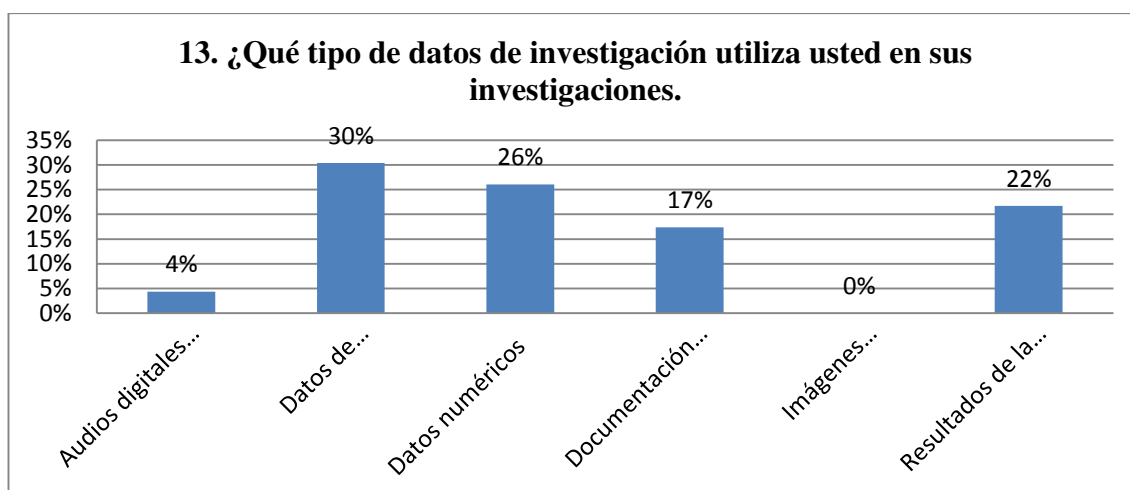
La utilización de la técnica observacional, en el caso de la ciencia de la salud, se debe principalmente porque los estudios se realizan sobre grupos de personas en los que se pueden observar los síntomas de una enfermedad, o se pueden observar los cambios de carácter físico. Por otro lado, he investigaciones en ciencias de la salud, la encuesta a través de un cuestionario permite conocer una determinada situación a partir de fuentes primarias.

Tabla 31 Distribución de frecuencias: tipo de datos de investigación que utilizan los investigadores de CS de la UNMSM

13. ¿Qué tipo de datos de investigación utiliza usted en sus investigaciones?

Respuesta	fi	hi
Audios digitales (archivos sonoros, entrevistas)	1	4%
Datos de encuestas	7	30%
Datos numéricos	6	26%
Documentación diversa	4	17%
Imágenes digitales (fotografía, diagramas)	0	0%
Resultados de la medida de instrumentos	5	22%
Total	23	100%

Gráfico 32 Frecuencia relativa: tipo de datos de investigación que utilizan los investigadores de CS de la UNMSM



Se determinó que los investigadores utilizan varios tipos de datos a la vez. Así tenemos, que el 30% (7 investigadores) utiliza datos de encuestas. El 26% (6 investigadores) utilizan datos numéricos. El 17% (4 investigadores) utiliza

documentación diversa. Y ninguno utiliza imágenes digitales (fotografía, diagramas etc.).

Esto se debe a que las encuestas son una técnica relativamente económica y fácil de aplicar, mientras que para manejar imágenes digitales se requiere de equipos más sofisticados y de conocimiento sobre los mismos.

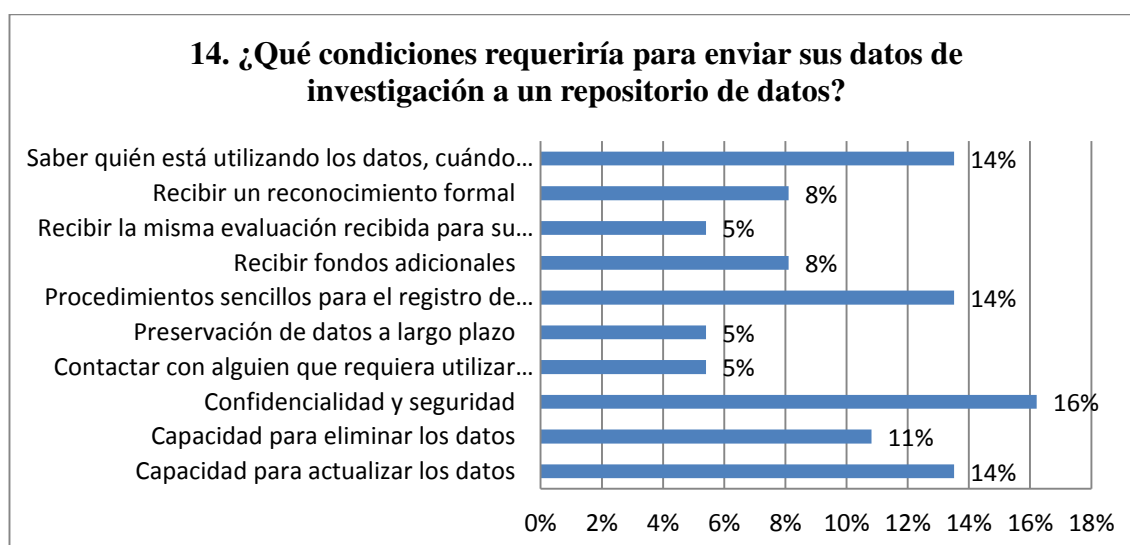
Bloque IV: almacenamiento de datos de investigación

Tabla 32 Distribución de frecuencias: condiciones requeridas por los investigadores de CS de la UNMSM para enviar sus datos de investigación a un repositorio

14. ¿Qué condiciones requeriría para enviar sus datos de investigación a un repositorio de datos?

Respuesta	fi	hi
Capacidad para actualizar los datos	5	14%
Capacidad para eliminar los datos	4	11%
Confidencialidad y seguridad	6	16%
Contactar con alguien que requiera utilizar mis datos	2	5%
Preservación de datos a largo plazo	2	5%
Procedimientos sencillos para el registro de los datos	5	14%
Recibir fondos adicionales	3	8%
Recibir la misma evaluación recibida para su publicación en una revista	2	5%
Recibir un reconocimiento formal	3	8%
Saber quién está utilizando los datos, cuándo y para qué fin	5	14%
Total	37	100%

Gráfico 33 Frecuencia relativa: condiciones requeridas por los investigadores de CS de la UNMSM para enviar sus datos de investigación a un repositorio



Se determinó que los investigadores requieren más de una condición para enviar sus datos de investigación a un repositorio de datos. Así tenemos, que el 16% (6 investigadores) requiere de confidencialidad y seguridad para enviar sus datos de investigación a un repositorio de datos. El 14% (5 investigadores) considera que una condición para enviar sus datos de investigación a un repositorio de datos es la capacidad para actualizar los datos. De igual manera otro 14% (5 investigadores) requieren de procedimientos sencillos para el registro de los datos. Sólo un 8% (3 investigadores) considera que la condición es recibir fondos adicionales para enviar sus datos de investigación. Y otro 8% considera que le bastaría recibir un reconocimiento formal.

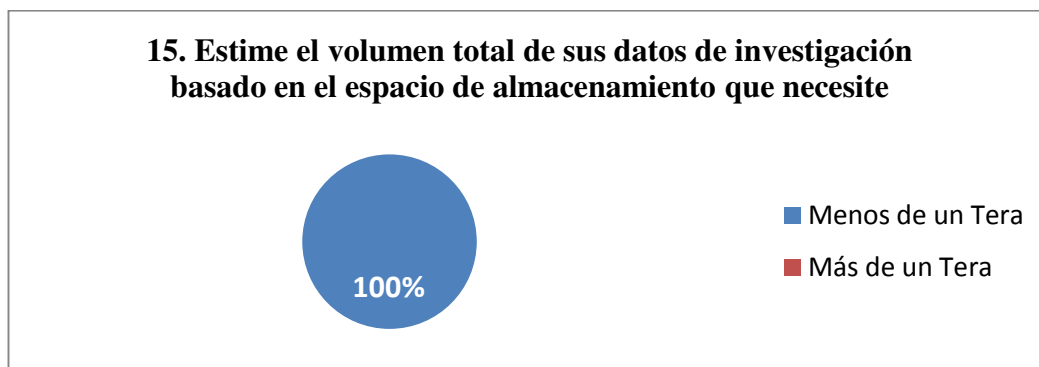
Es evidente, la necesidad de contar con un repositorio de datos seguro y confiable, el tema es, que actualmente, las plataformas que utilizan los repositorios, principalmente universitarios, cumplan con los protocolos de seguridad, confidencialidad, actualización, prevención, preservación y publicación. En ese sentido, se debería contar con un programa de difusión, comunicación y capacitación sobre los beneficios de registrar los datos de investigación en un repositorio de datos.

Tabla 33 Distribución de frecuencias: espacio de almacenamiento que necesita los investigadores de CS de la UNMSM

15. Estime el volumen total de sus datos de investigación basado en el espacio de almacenamiento que necesite (el promedio estimado por investigación).

Respuesta	fi	hi
Menos de un Tera	9	100%
Más de un Tera	0	0%
Total	9	100 %

Gráfico 34 Frecuencia relativa: espacio de almacenamiento que necesita los investigadores de CS de la UNMSM



Se determinó que el 100% (9 investigadores) requiere como espacio de almacenamiento por investigación, menos de un Tera.

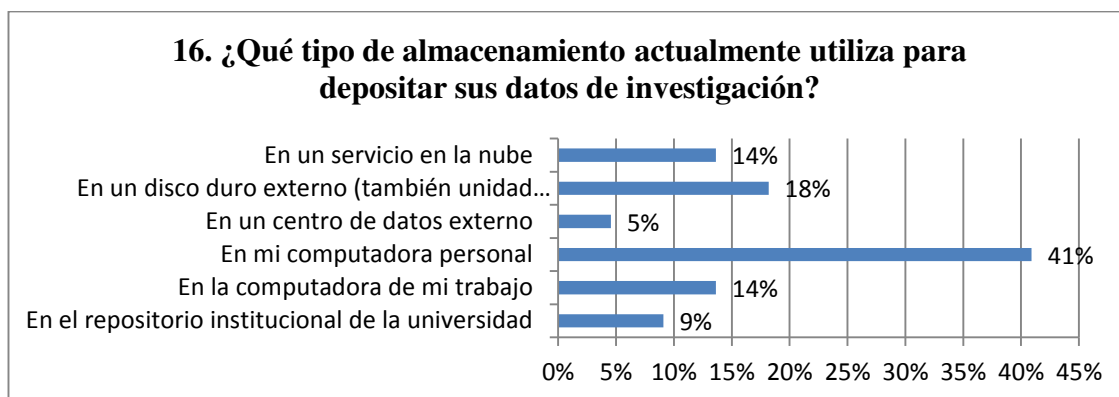
Esto se debe, a que los investigadores no utilizan archivos pesados como audios digitales (archivos sonoros, entrevistas etc.) o Imágenes digitales (fotografía, diagramas, etc.). Además, que la tendencia para conservar archivos o datos de investigación es comprimir los mismos.

Tabla 34 Distribución de frecuencias: tipo de almacenamiento actualmente utiliza el investigador de CS de la UNMSM

16. ¿Qué tipo de almacenamiento actualmente utiliza para depositar sus datos de investigación?

Respuesta	fi	hi
En el repositorio institucional de la universidad	2	9%
En la computadora de mi trabajo	3	14%
En mi computadora personal	9	41%
En un centro de datos externo	1	5%
En un disco duro externo (también unidad USB)	4	18%
En un servicio en la nube	3	14%
Total	22	100%

Gráfico 35 Frecuencia relativa: tipo de almacenamiento que actualmente utiliza el investigador de CS de la UNMSM



Se determinó que los investigadores utilizan más de un tipo de almacenamiento. Sin embargo, el 41% (9 investigadores) utiliza su computadora personal. El 18% (4 investigadores) utiliza un disco duro externo, así como, unidades USB. El 14% (3 investigadores) utiliza la computadora de su trabajo. Otro 14% (3 investigadores) utiliza el servicio de la nube. Y el 9% (2 investigadores) utiliza el repositorio institucional de la Universidad.

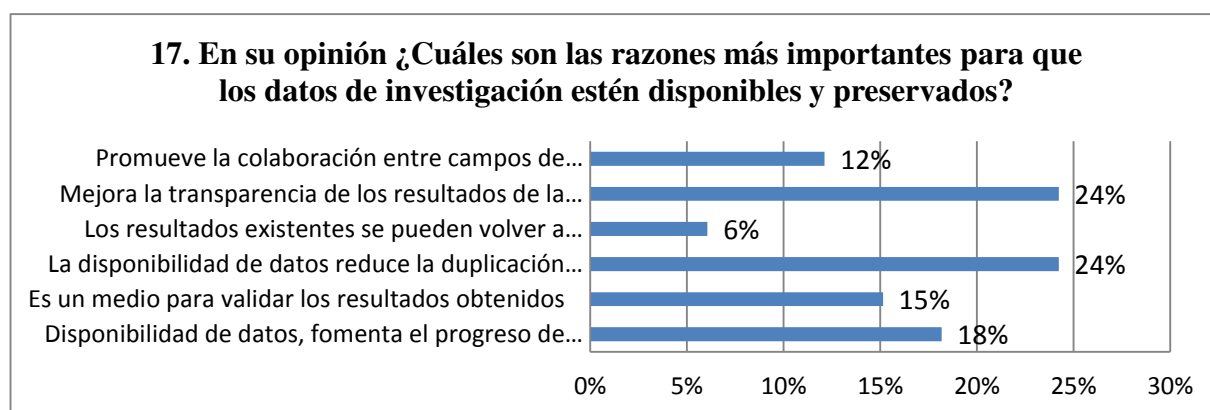
Queda claro, que la oferta de tecnología en tipos de almacenamiento es variada, esta situación le permite al investigador escoger el tipo de almacenamiento que le conviene. Sin embargo, no utiliza el repositorio institucional debido a que no existe en su institución o no tiene conocimiento de la importancia del mismo para poder compartir, reutilizar y asegurar sus datos de investigación en una plataforma confiable.

Tabla 35 Distribución de frecuencias: razones más importantes para los investigadores de CS de la UNMSM para que sus datos de investigación estén disponibles y preservados

17. En su opinión ¿Cuáles son las razones más importantes para que los datos de investigación estén disponibles y preservados?

Respuesta	fi	hi
Disponibilidad de datos, fomenta el progreso de la ciencia (nueva investigación se basa en el conocimiento preexistente)	6	18%
Es un medio para validar los resultados obtenidos	5	15%
La disponibilidad de datos reduce la duplicación de esfuerzos de investigación	8	24%
Los resultados existentes se pueden volver a examinar	2	6%
Mejora la transparencia de los resultados de la investigación	8	24%
Promueve la colaboración entre campos de diferentes disciplinas	4	12%
Total	33	100%

Gráfico 36 Frecuencia relativa: de las razones más importantes para los investigadores de CS de la UNMSM para que sus datos de investigación estén disponibles y preservados



Se determinó que los investigadores tienen más de una razón importante para que los datos de investigación estén disponibles y preservados. El 24% (8 investigadores) considera que la disponibilidad de datos reduce la duplicación de esfuerzos de investigación. Otro 24% (8 investigadores) considera que una razón importante por la cual los datos de investigación deberían estar disponibles y preservados es la transparencia de los resultados de la investigación. El 18% (6 investigadores) considera que la disponibilidad de datos, fomenta el progreso de la ciencia. El 15% (5 investigadores) considera que si los datos de investigación están disponibles y preservados es un medio para validar los resultados obtenidos. El 12%

(4 investigadores) entiende que la disponibilidad y preservación de los datos promueve la colaboración entre campos de diferentes disciplinas. Y sólo el 6% (2 investigadores) visualiza que teniendo disponibles y preservados los datos, los resultados existentes se pueden volver a examinar.

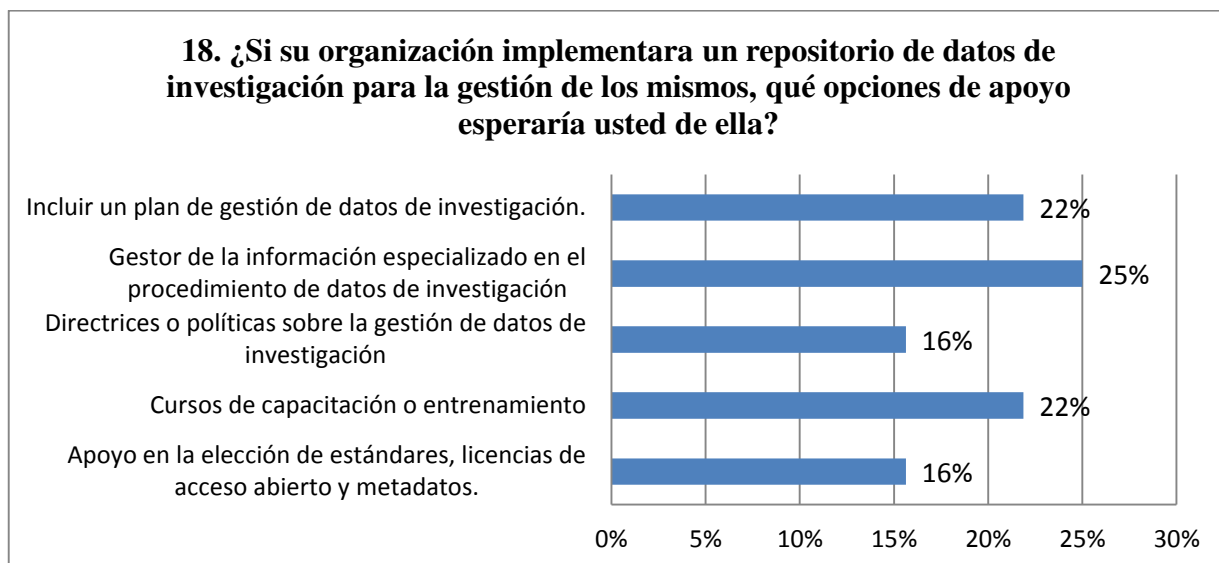
La disponibilidad de datos de investigación además de evitar un doble esfuerzo en su búsqueda y promover la transparencia, fomenta y contribuye a la investigación, lo que tiene como consecuencia la generación de nuevo conocimiento. Un conocimiento que tendría que estar disponible y preservado y a su vez constituirse en una cadena sin fin de generación, producción y desarrollo de conocimientos.

Tabla 36 Distribución de frecuencias: opciones de apoyo que el investigador de CS de la UNMSM esperaría de su institución si implementará un repositorio

18. ¿Si su organización implementara un repositorio de datos de investigación para la gestión de los mismos, qué opciones de apoyo esperaría usted de ella?

Respuesta	fi	hi
Apoyo en la elección de estándares, licencias de acceso abierto y metadatos.	5	16%
Cursos de capacitación o entrenamiento	7	22%
Directrices o políticas sobre la gestión de datos de investigación	5	16%
Gestor de la información especializado en el procedimiento de datos de inv.	8	25%
Incluir un plan de gestión de datos de investigación.	7	22%
Total	32	100 %

Gráfico 37 Frecuencia relativa: opciones de apoyo que el investigador de CS de la UNMSM esperaría de su institución si implementará un repositorio



Se determinó que los investigadores esperan más de un tipo de apoyo de parte de su organización para compartir sus datos de investigación a través de un repositorio. El 25% (8 investigadores) esperaría que el manejo de sus datos de investigación fuera un gestor de la información especializado en el procedimiento de datos de investigación. El 22% (7 investigadores) esperaría como apoyo cursos de capacitación o entrenamiento. Otro 22% (7 investigadores) esperaría que incluyan un plan de gestión de datos de investigación. El 16% (5 investigadores) esperaría apoyo en la elección de estándares y licencias de acceso abierto en metadatos. Y otro 16% (5 investigadores) esperaría apoyo en directrices o políticas sobre la gestión de datos de investigación.

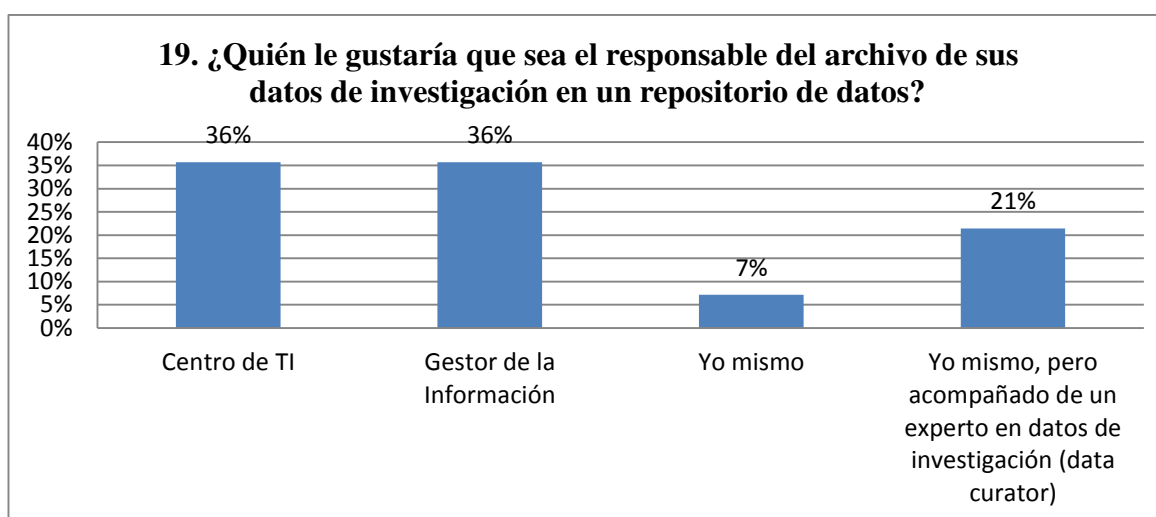
En este contexto, cualquier tipo de apoyo en el campo de la investigación será muy apreciado por los investigadores. En el Perú el número de investigaciones que se realizan tanto en universidades privadas como en universidades públicas es todavía mínima, en comparación con otros países de la región.

Tabla 37 Distribución de frecuencias: preferencia del responsable para hacerse cargo del archivo de datos de investigación de los investigadores de CS de la UNMSM

19. ¿Quién le gustaría que sea el responsable del archivo de sus datos de investigación en un repositorio de datos?

Respuesta	fi	hi
Centro de TI	5	36%
Gestor de la Información	5	36%
Yo mismo	1	7%
Yo mismo, pero acompañado de un experto en datos de investigación (data curator)	3	21%
Total	14	100%

Gráfico 38 Frecuencia relativa: preferencia del responsable para hacerse cargo del archivo de datos de investigación de los investigadores de CS de la UNMSM



Se determinó que al 36% (5 investigadores) le gustaría que el responsable del archivo de sus datos sea el Centro de TI. A otro 36% (5 investigadores) le gustaría que el responsable del archivo de sus datos sea un Gestor de la Información. A un 21% (3 investigadores) le gustaría ser el mismo el responsable del archivo acompañado por un experto en datos de investigación (data curator), y sólo el 7% (1 investigador) preferiría ser el mismo el responsable del archivo de sus datos de investigación.

Se puede observar que la mayoría de investigadores están dispuestos a colocar sus de datos de investigación, en un repositorio de datos, siempre y cuando le garanticen la presencia de un experto o de un gestor de información.

5.2. Análisis e Interpretación de los resultados UPC

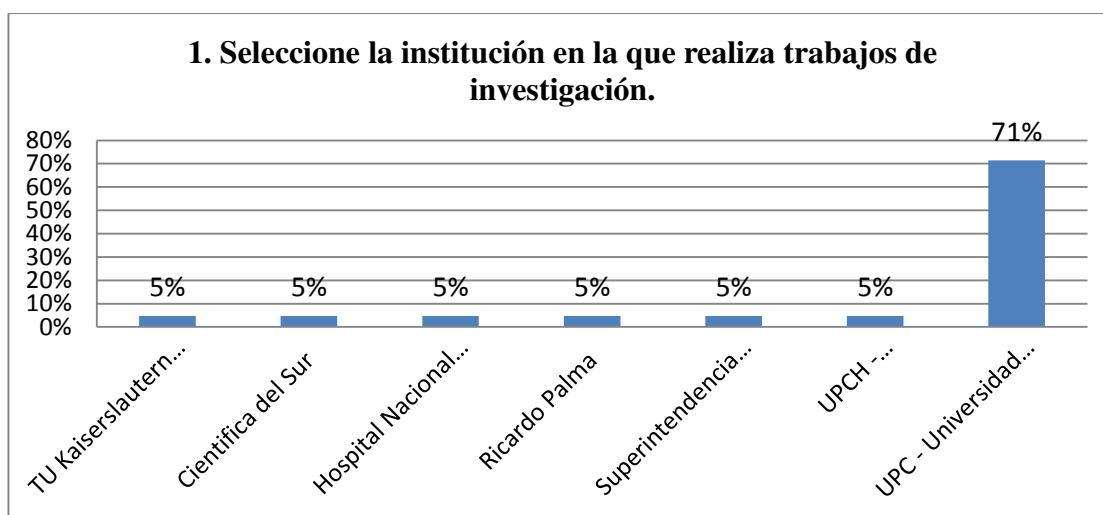
Bloque I: información general

Tabla 38 Distribución de frecuencias: instituciones donde los investigadores de CS de la UPC realizan sus trabajos de investigación

1. Seleccione la institución en la que realiza trabajos de investigación.

Respuesta	fi	hi
TU Kaiserslautern (Alemania)	1	5%
Cientifica del Sur	1	5%
Hospital Nacional 2 de mayo	1	5%
Ricardo Palma	1	5%
Superintendencia Nacional de Salud	1	5%
UPCH - Universidad Peruana Cayetano Heredia	1	5%
UPC - Universidad Peruana de Ciencias Aplicadas	15	71%
TOTAL DE RESPUESTAS	21	100%

Gráfico 39 Frecuencia relativa: instituciones donde los investigadores de la UPC realizan sus trabajos de investigación



Se determinó que los investigadores utilizan más de una institución para realizar trabajos de investigación. El 71% (15 investigadores) realizan sus investigaciones en la Universidad Peruana de Ciencias Aplicadas, UPC. Y el 29% (6 investigadores) realizan sus investigaciones en otras instituciones privadas. Esto se

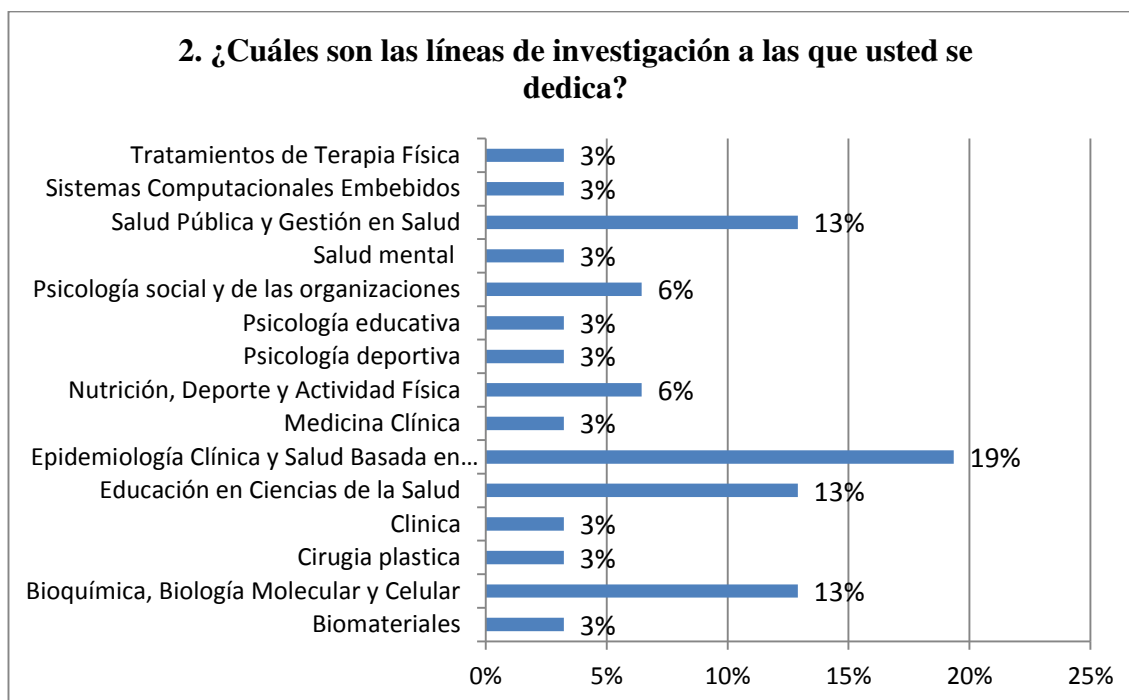
debe principalmente a que la investigación en la UPC es un lineamiento estratégico y por lo tanto tiene todo el apoyo de la alta dirección de la institución, con este fin los recursos disponibles para investigación van incrementándose año a año.

Tabla 39 Distribución de frecuencias: líneas de investigación a las que se dedican los investigadores de CS de la UPC

2. ¿Cuáles son las líneas de investigación a las que usted se dedica?

Respuesta	fi	hi
Biomateriales	1	3%
Bioquímica, Biología Molecular y Celular	4	13%
Cirugía plastica	1	3%
Clinica	1	3%
Educación en Ciencias de la Salud	4	13%
Epidemiología Clínica y Salud Basada en Evidencias	6	19%
Medicina Clínica	1	3%
Nutrición, Deporte y Actividad Física	2	6%
Psicología deportiva	1	3%
Psicología educativa	1	3%
Psicología social y de las organizaciones	2	6%
Salud mental	1	3%
Salud Pública y Gestión en Salud	4	13%
Sistemas Computacionales Embebidos	1	3%
Tratamientos de Terapia Física	1	3%
TOTAL DE RESPUESTAS	31	100%

Gráfico 40 Frecuencia relativa: líneas de investigación a las que se dedican los investigadores de la UPC



Se determinó que las líneas de investigación a las que se dedican los investigadores es más de una. El 19% (6 investigadores) tiene como línea de investigación la Epidemiología clínica y salud basada en evidencias. Un 13% (4 investigadores) tiene como línea de investigación la Bioquímica, Biología Molecular y Celular. Otro 13% (4 investigadores) tiene como línea de investigación la Educación en Ciencias de la Salud. Y otro 13% (4 investigadores) tiene como línea de investigación de la Salud Pública y Gestión en Salud. Y el 42% (13 investigadores) tiene como líneas de investigación los Biomateriales, Cirugía Plástica, Clínica, Medicina Clínica, Nutrición, Deporte y Actividad Física, Psicología Deportiva, Psicología Educativa, Psicología Social y de las Organizaciones, Sistemas Computacionales Embebidos y Tratamiento de Terapia Física.

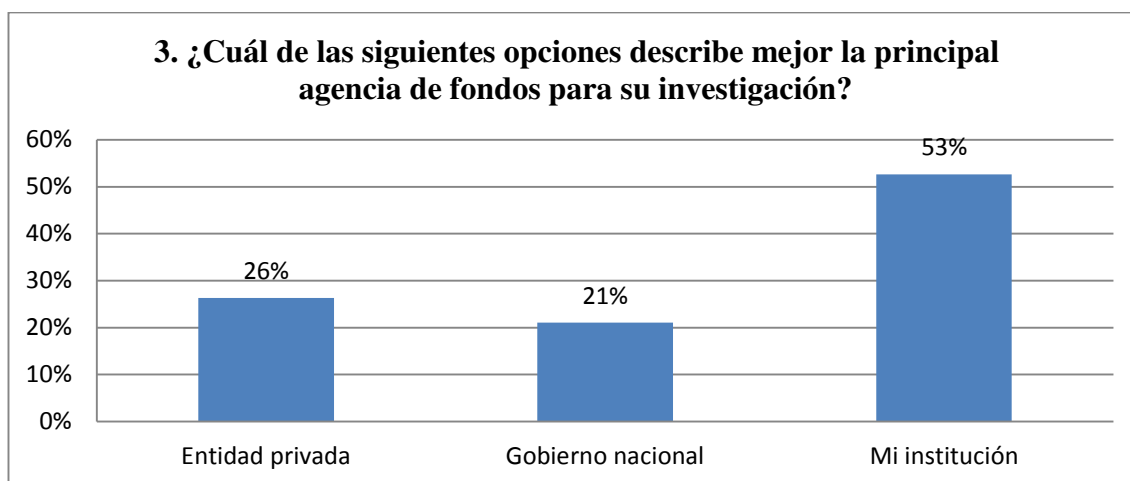
La investigación y los investigadores en la UPC vienen teniendo cada vez más atención, bajo la propuesta de la alta dirección de la Universidad de que la investigación es una fuente de generación de conocimiento, que debe ser apoyada de manera eficiente, una de estas maneras es proveerle al investigador de un conjunto de herramientas que facilite su trabajo.

Tabla 40 Distribución de frecuencias: agencias que contribuyen con fondos a la investigación en UPC

3. ¿Cuál de las siguientes opciones describe mejor la principal agencia de fondos para su investigación?

Respuesta	fi	hi
Entidad privada	5	26%
Gobierno nacional	4	21%
Mi institución	10	53%
TOTAL DE RESPUESTAS	19	100%

Gráfico 41 Frecuencia relativa: agencias que contribuyen con fondos a la investigación en UNMSM



Se determinó que los investigadores son financiados por más de una agencia. El 53% (10 investigadores) son financiados por la UPC. El 26% (5 investigadores) también reciben fondos de entidades privadas. Y el 21% (4 investigadores) reciben fondos para sus investigaciones del gobierno nacional.

En lo que respecta al primer caso, la UPC tiene a la investigación, como una línea estratégica institucional, por lo que los fondos para investigación vienen incrementándose año a año. Respecto a los fondos que provienen de entidades privadas y del gobierno nacional todavía son limitados.

Bloque II: uso de datos de investigación

Tabla 41 Distribución de frecuencias: uso que le da el investigador de CS de la UPC

4. ¿Qué hace usted con sus datos de investigación?

Respuesta	fi	hi
Los conserva	18	100%
No los conserva	0	0%

Gráfico 42 Frecuencia relativa: uso que le da el investigador de la UPC a sus de datos de investigación

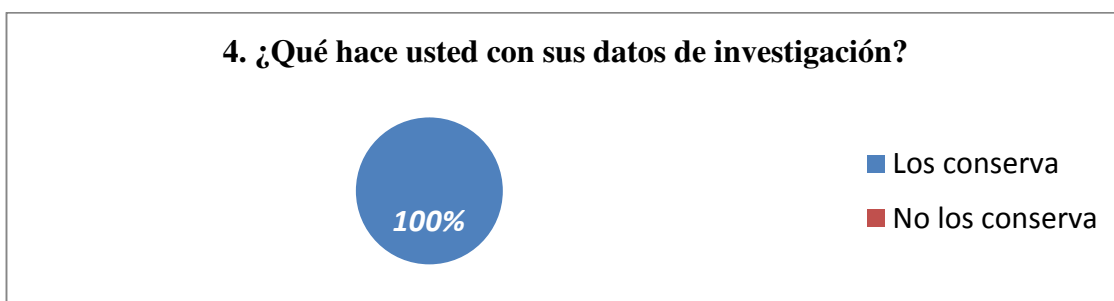
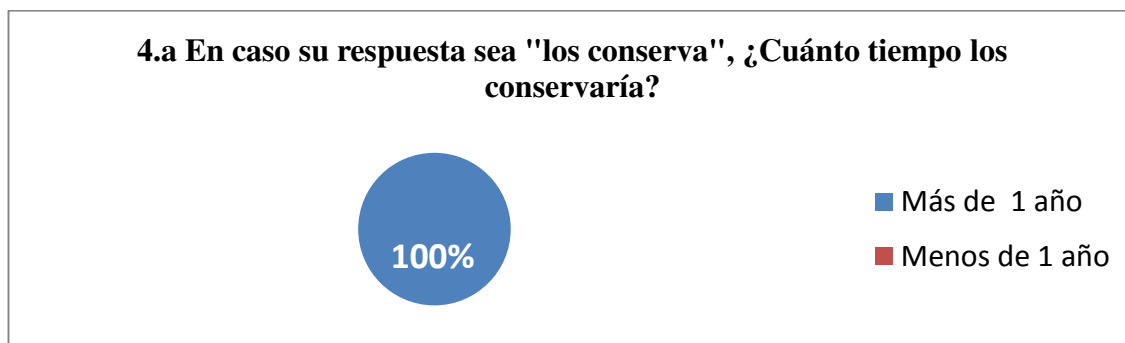


Tabla 42 Distribución de frecuencias: tiempo que conserva el investigador de CS de la UPC sus datos de investigación

4.a En caso su respuesta sea "Los conserva", ¿Cuánto tiempo los conservaría?

Respuesta	fi	hi
Más de 1 año	16	100%
Menos de 1 año	0	0%

Gráfico 43 Frecuencia relativa: tiempo que conserva el investigador de la UPC sus datos de investigación



Se determinó que el 100% de los investigadores conserva sus datos de investigación. Respecto al tiempo que el investigador conserva sus datos de investigación, el 100% (16 investigadores) lo hace por más de un año.

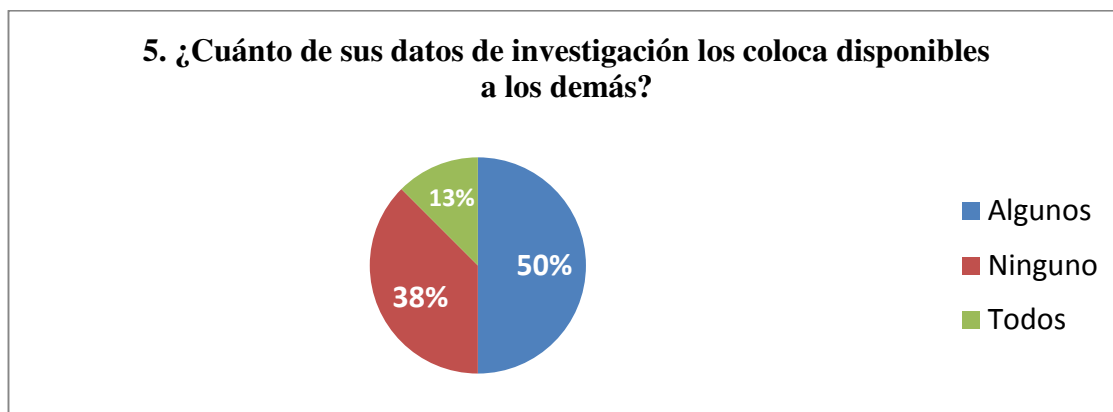
El conservar los datos de investigación se debe al sentido de previsión para el uso de los mismos en un futuro cercano o al sentido de protección de la propiedad intelectual.

Tabla 43 Distribución de frecuencias: cantidad de datos de investigación que el investigador de CS de la UPC coloca disponibles a los demás

5. ¿Cuánto de sus datos de investigación los coloca disponibles a los demás?

Respuesta	fi	hi
Algunos	8	50%
Ninguno	6	38%
Todos	2	13%
Total	16	100%

Gráfico 44 Frecuencia relativa: cantidad de datos de investigación que el investigador de la UPC pone a disponibilidad de los demás



Se determinó que el 13% (2 investigadores) ponen todos sus datos de investigación a disposición de los demás. El 50% (8 investigadores) ponen algunos de sus datos de investigación a disposición de los demás. Y el 38% (6 investigadores) no pone ninguno.

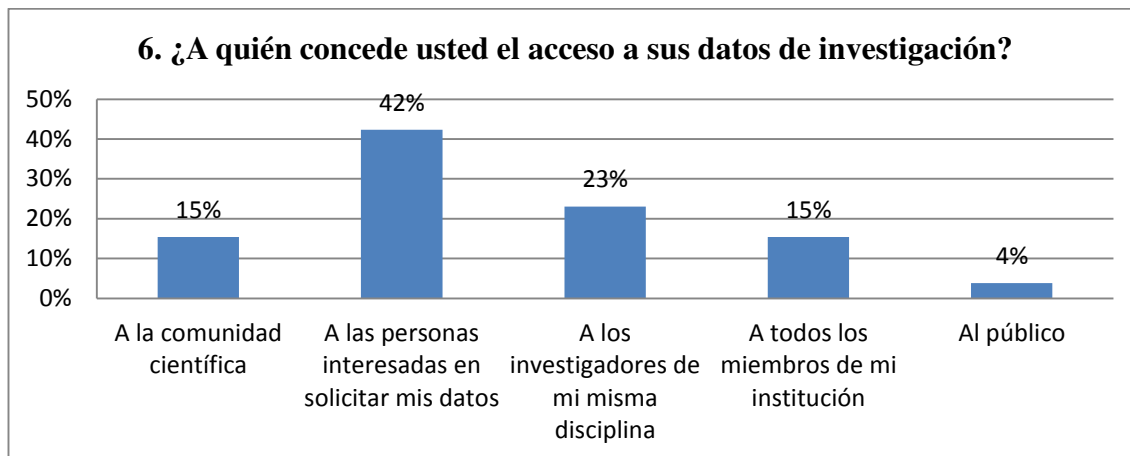
El no colocar sus datos de investigación a disposición de los demás, se debe a que no hay un medio confiable en el que puedan interactuar e intercambiar datos de investigación entre investigadores.

Tabla 44 Distribución de frecuencias: usuarios que tienen acceso a los datos de investigación de los investigadores de CS de la UPC

6. ¿A quién concede usted el acceso a sus datos de investigación?

Respuesta	fi	hi
A la comunidad científica	4	15%
A las personas interesadas en solicitar mis datos	11	42%
A los investigadores de mi misma disciplina	6	23%
A todos los miembros de mi institución	4	15%
Al público	1	4%
Total	26	100%

Gráfico 45 Frecuencia relativa: usuarios que tiene acceso a los datos de investigación de los investigadores de CS de la UPC



Se determinó que los investigadores dan acceso a sus datos de investigación a más de una entidad. El 42% (11 investigadores) le dan acceso a sus datos de investigación a las personas interesadas que se lo solicitan. El 23% (6 investigadores) le da acceso a sus datos de investigación a los investigadores de su misma disciplina. El 15% (4 investigadores) le da acceso a la comunidad y científica. Y otro 15% (4 investigadores) le da acceso a todos los miembros de su institución.

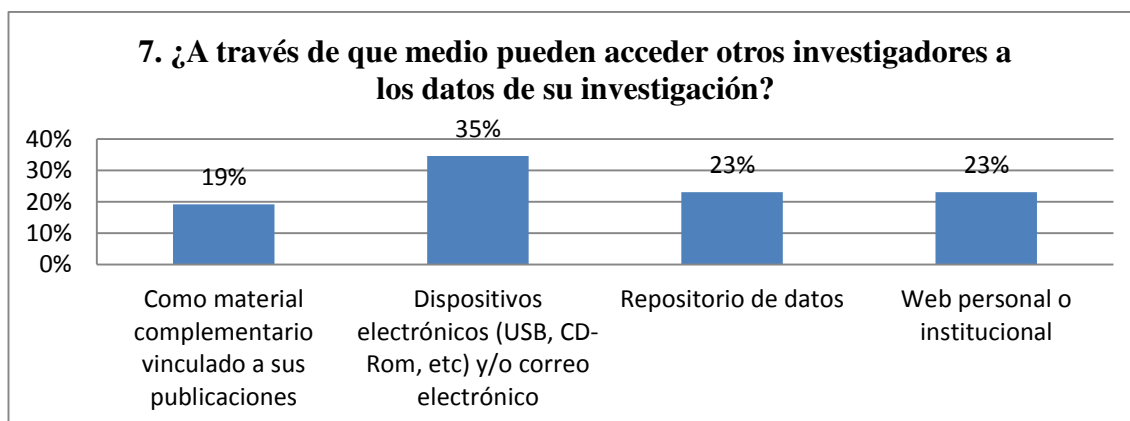
Se observa, qué hay una buena disposición, de parte de los investigadores, para dar acceso a sus datos de investigación. Por otro lado, los investigadores prefieren interactuar por lo menos una vez con aquellas personas que le solicitan sus datos de investigación y también con los investigadores de su institución que conocen y que comparten la misma línea de investigación.

Tabla 45 Distribución de frecuencias: medios a través de los cuales se accede a los datos de investigación de los investigadores de CS de la UPC

7. ¿A través de que medio pueden acceder otros investigadores a los datos de su investigación?

Respuesta	fi	hi
Como material complementario vinculado a sus publicaciones	5	19%
Dispositivos electrónicos (USB, CD-Rom, etc) y/o correo electrónico	9	35%
Repositorio de datos	6	23%
Web personal o institucional	6	23%
Total	26	100%

Gráfico 46 Frecuencia relativa: medios a través de los cuales se accede a los datos de investigación de los investigadores de CS de la UPC



Se determinó que los investigadores pueden dar acceso a sus datos de investigación a otros investigadores a través de más de un medio. El 35% (9 investigadores) dar acceso a sus datos de investigación a través de dispositivos electrónicos (USB, CD-Rom, etc.) y/o correo electrónico. El 23% (6 investigadores) da acceso a sus datos de investigación a través de repositorio de datos. Otro 23% (6 investigadores) lo hace a través de la web personal o institucional. Y un 19% (5 investigadores) da acceso a sus datos de investigación a otros investigadores como material complementario vinculado a su publicación.

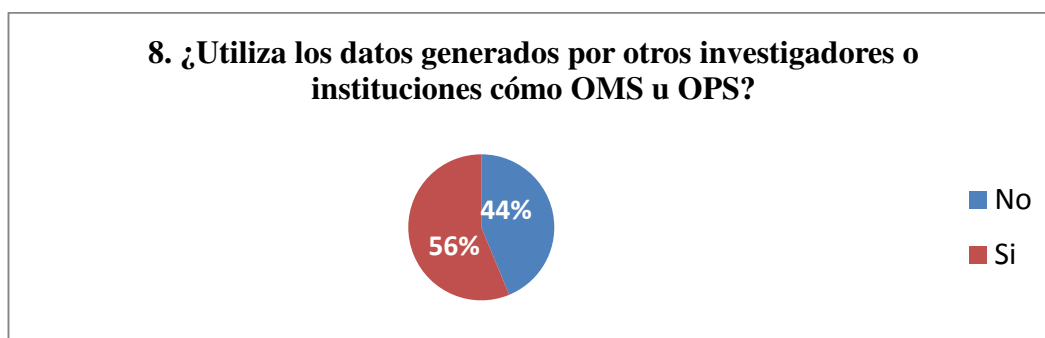
Se observa que los investigadores utilizan aquellos medios o dispositivos electrónicos (USB, CD-Rom, etc.) y/o correo electrónico que están a su alcance, fáciles de manipular y que a su vez son económicos o gratuitos.

Tabla 46 Distribución de frecuencias: utilización de datos generados por otros investigadores o instituciones, realizados por los investigadores de CS de la UPC

8. ¿Utiliza los datos generados por otros investigadores o instituciones cómo OMS u OPS?

Respuesta	fi	hi
No	7	44%
Si	9	56%
Total	16	

Gráfico 47 Frecuencia relativa: utilización de datos generados por otros investigadores o instituciones, realizados por los investigadores de CS de la UPC



Se determinó que el 56% (9 investigadores) utiliza los datos generados por otros investigadores o instituciones como la OMS u OPS. Y el 44% (7 investigadores) no utiliza datos generados por otros investigadores.

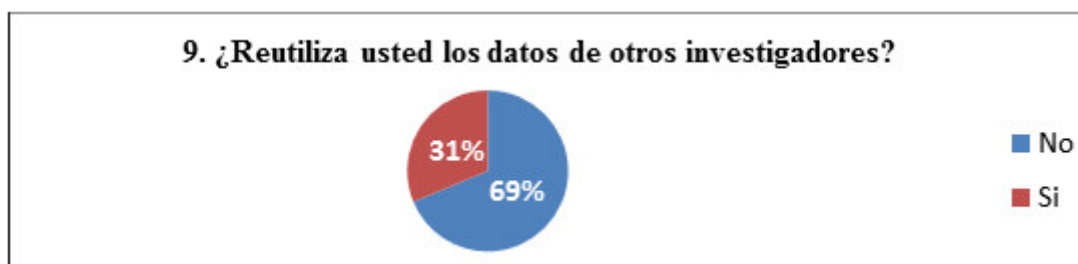
El no usar datos generados por otros investigadores o instituciones como la OMS u OPS, se debe a la falta de información sobre la posibilidad de utilizar los datos generados por otros investigadores sin contraer ningún compromiso económico, administrativo o académico.

Tabla 47 Distribución de frecuencias: reutilización de datos generados por otros investigadores o instituciones, realizados por los investigadores de CS de la UPC

9. ¿Reutiliza usted los datos de otros investigadores?

Respuesta	fi	hi
No	11	69%
Si	5	31%
Total	16	

Gráfico 48 Frecuencia relativa: reutilización de datos generados por otros investigadores o instituciones, por los investigadores de CS de la UPC



Se determinó que el 69% (11 investigadores) no reutilizan los datos de otros investigadores. Sólo el 31% (5 investigadores) lo hace.

El no reutilizar datos de investigación de otros investigadores se debe a la falta de información sobre los beneficios de la reutilización, toda vez que significa, ahorro de recursos y control de esfuerzos.

Tabla 48 Distribución de frecuencias: tipo de licencia que aplican los investigadores de CS de la UPC

9.a. En caso su respuesta sea afirmativa, seleccione el tipo de licencia que aplicaría.

Respuesta	fi	hi
Acuerdos de cooperación	1	14%
Creative Commons	1	14%
Ninguno	1	14%
No sé	2	29%
Open Data License Attribution	2	29%
Total	7	

Gráfico 49 Frecuencia relativa: tipo de licencia que aplican los investigadores de CS de la UPC



Se determinó que los investigadores que reutilizan datos de investigación de otros investigadores, seleccionaría más de un tipo de licencia. El 29% (2 investigadores) seleccionaría la licencia Open Data License Attribution. Otro 29% (2 investigadores) no sabe qué tipo de licencia seleccionar. El 14% (1 investigador) seleccionaría como licencia los acuerdos de cooperación. Otro 14% (1 investigador) seleccionaría la licencia Creative Commons. Y por último, un 14% (1 investigador) no seleccionaría ningún tipo de licencia.

Se debe tomar en cuenta, que las licencias para el uso de recursos digitales, que permiten publicar las investigaciones, tiene una ventaja importante que es la protección de la propiedad intelectual y el derecho de autor, así como, la oportunidad de que el autor tenga visibilidad.

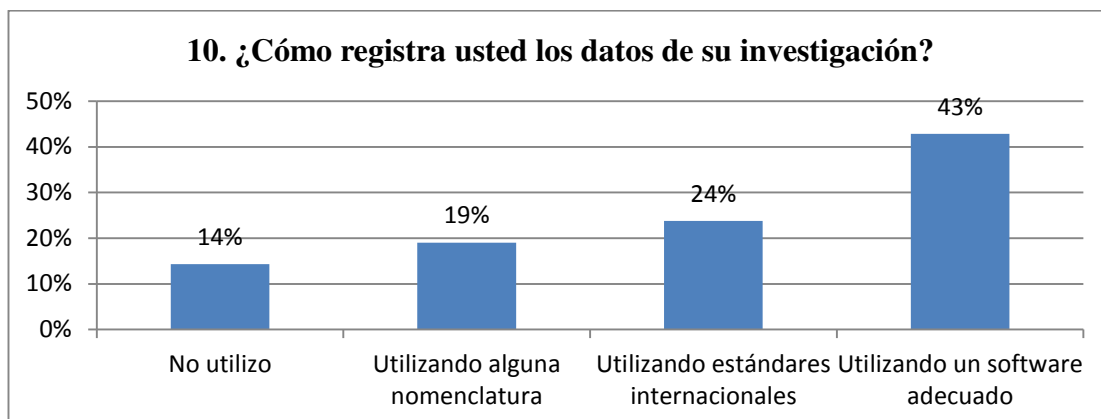
Bloque III: organización de datos de investigación

Tabla 49 Distribución de frecuencias: sobre como registra sus datos de investigación los investigadores de CS de la UPC

10. ¿Cómo registra usted los datos de su investigación?

Respuesta	fi	hi
No utilizo	3	14%
Utilizando alguna nomenclatura	4	19%
Utilizando estándares internacionales	5	24%
Utilizando un software adecuado	9	43%
Total	21	100%

Gráfico 50 Frecuencia relativa: sobre como registra sus datos de investigación los investigadores de CS de la UPC



Se determinó que los investigadores tienen más de una forma de registrar sus datos de investigación. El 43% (9 investigadores) registra sus datos de investigación utilizando un software adecuado. El 24% (5 investigadores) registra sus datos de investigación utilizando estándares internacionales. El 19% (4 investigadores) lo hace utilizando alguna nomenclatura. Y el 14% (3 investigadores) no registra sus datos de investigación.

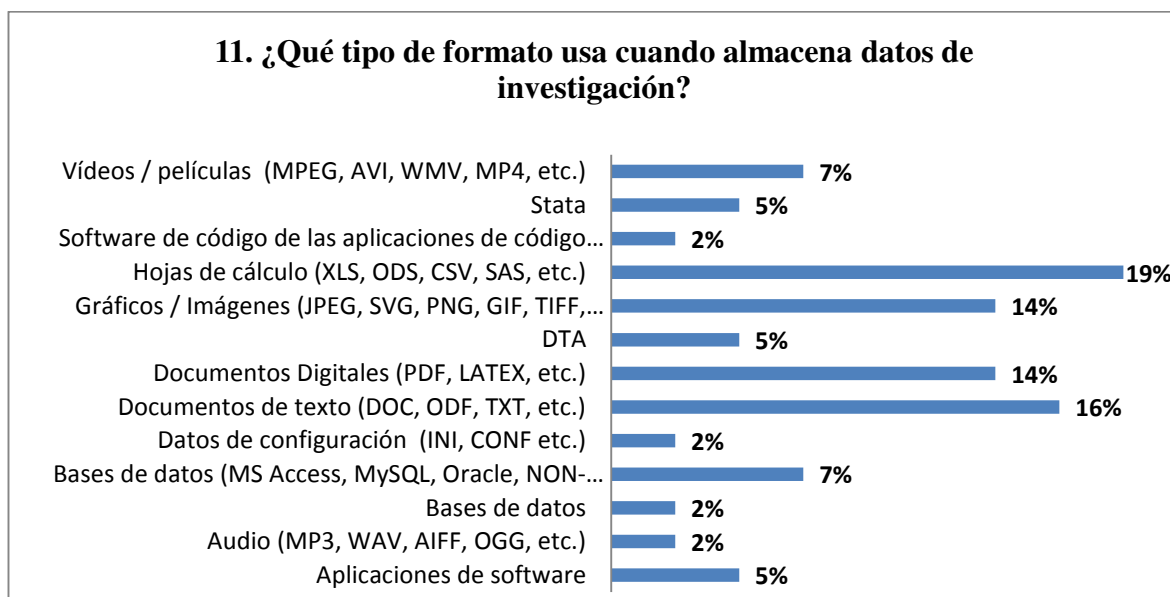
Se debe tener presente que el proceso de registrar los datos de investigación, es parte sustancial de la investigación misma, de tal forma, que durante la investigación que se generaron datos de investigación o en una nueva investigación, estos puedan ser utilizados, más de una vez, siempre y cuando estén debidamente registrados.

Tabla 50 Distribución de frecuencias: tipo de formato que usa el investigador de CS de la UPC para almacenar sus datos de investigación

11. ¿Qué tipo de formato usa cuando almacena datos de investigación?

Respuesta	fi	hi
Aplicaciones de software	2	5%
Audio (MP3, WAV, AIFF, OGG, etc.)	1	2%
Bases de datos	1	2%
Bases de datos (MS Access, MySQL, Oracle, NON-SQL, etc.)	3	7%
Datos de configuración (INI, CONF etc.)	1	2%
Documentos de texto (DOC, ODF, TXT, etc.)	7	16%
Documentos Digitales (PDF, LATEX, etc.)	6	14%
DTA	2	5%
Gráficos / Imágenes (JPEG, SVG, PNG, GIF, TIFF, etc.)	6	14%
Hojas de cálculo (XLS, ODS, CSV, SAS, etc.)	8	19%
Software de código de las aplicaciones de código (CSS, JavaScript, Java, etc.)	1	2%
Stata	2	5%
Videos / películas MPEG, AVI, WMV, MP4, etc.)	3	7%
Total	43	100%

Gráfico 51 Frecuencia relativa: tipo de formato que usa el investigador de CS de la UPC para almacenar sus datos de investigación



Se determinó que los investigadores utilizan más de un formato para almacenar sus datos de investigación. El 19% (8 investigadores) utiliza Hojas de cálculo (XLS, ODS, CSV, SAS, etc.). El 16% (7 investigadores) utiliza documentos de texto (DOC, ODF, TXT, etc.). Un 14% (6 investigadores) utiliza documentos digitales (PDF,

LATEX, etc.). Otro 14% (6 investigadores) utiliza gráficos / imágenes (JPEG, SVG, PNG, GIF, TIFF, etc.). Un 7% (3 investigadores) utiliza bases de datos (MS Access, MySQL, Oracle, NON-SQL, etc.). Y un 37% (16 investigadores) utilizan otro tipo de formato.

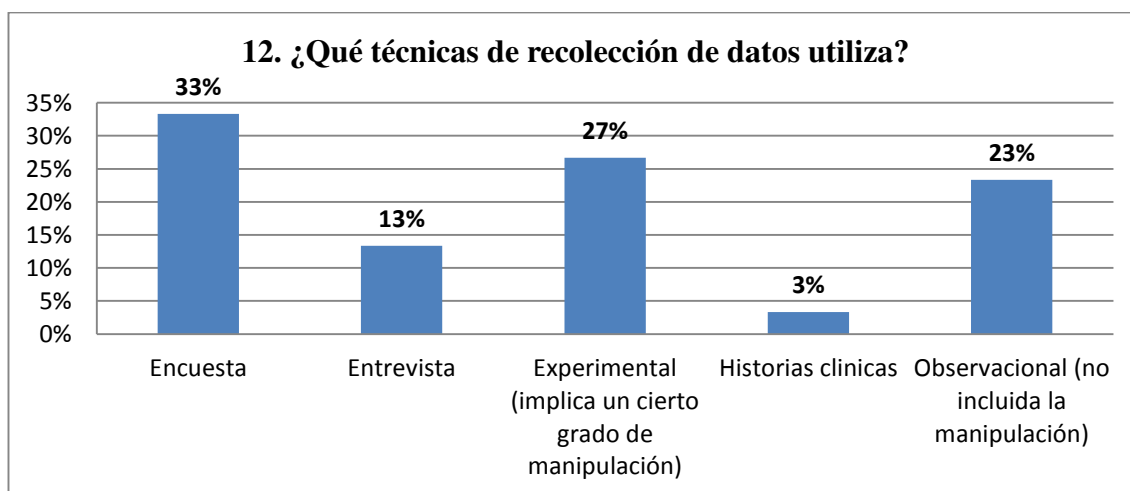
La mayoría de docentes investigadores busca organizar sus datos de una manera fácil y económica en softwares que están a su alcance como el Office de Microsoft, que están provistos de hojas de cálculo y de texto, siendo, algunos de ellos formatos adecuados para la reutilización de los datos, pero que no son utilizados como estándares en todas sus investigaciones para este fin,

Tabla 51 Distribución de frecuencias: técnicas de recolección de datos que utiliza el investigador de CS de la UPC

12. ¿Qué técnicas de recolección de datos utiliza?

Respuesta	fi	hi
Encuesta	10	33%
Entrevista	4	13%
Experimental (implica un cierto grado de manipulación)	8	27%
Historias clínicas	1	3%
Observacional (no incluida la manipulación)	7	23%
Total	30	100%

Gráfico 52 Frecuencia relativa: técnicas de recolección de datos que utiliza el investigador de CS de la UPC



Se determinó que los docentes investigadores de la UPC utilizan más de una técnica de recolección de datos. El 33% (10 investigadores) utiliza como técnica la encuesta. El 27% (8 investigadores) recolecta datos utilizando la técnica experimental. El 23% (7 investigadores) lo hace a través de la técnica observacional. El 13% (4 investigadores) utiliza la técnica de la entrevista. Y sólo el 3% (1 investigador) utiliza para recolectar sus datos las historias clínicas.

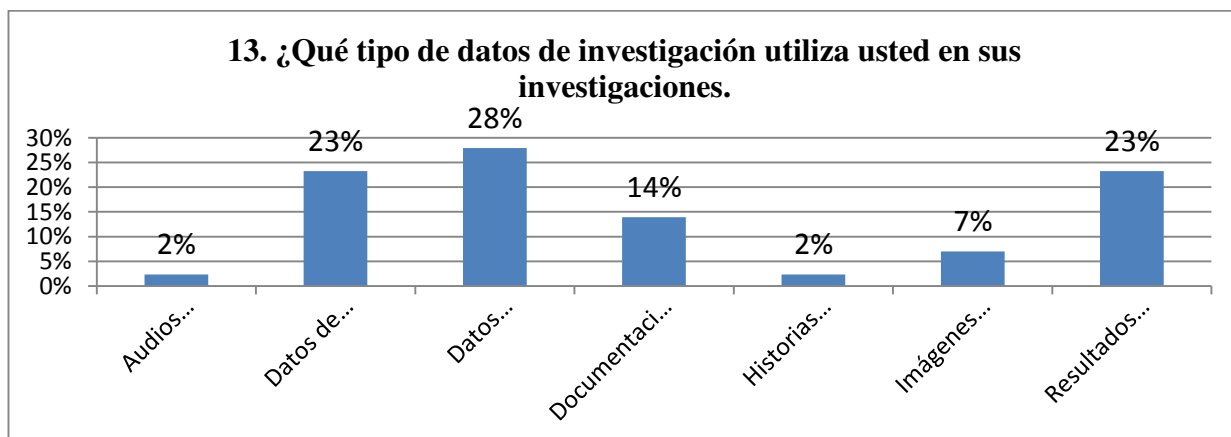
El uso de diferentes técnicas para la recolección de datos indica que los docentes investigadores tienen apoyo de parte de la Universidad para la organización y administración de sus datos.

Tabla 52 Distribución de frecuencias: tipo de datos de investigación que utilizan los investigadores de CS de la UPC

13. ¿Qué tipo de datos de investigación utiliza usted en sus investigaciones?

Respuesta	fi	hi
Audios digitales (archivos sonoros, entrevistas)	1	2%
Datos de encuestas	10	23%
Datos numéricos	12	28%
Documentación diversa	6	14%
Historias clínicas	1	2%
Imágenes digitales (fotografía, diagramas)	3	7%
Resultados de la medida de instrumentos	10	23%
Total	43	100%

Gráfico 53 Frecuencia relativa: tipo de datos de investigación que utilizan los investigadores de CS de la UPC



Se determinó que los docentes investigadores de la UPC utilizan más de un tipo de dato de investigación. El 28% (12 investigadores) utiliza datos numéricos. El 23% (10 investigadores) utiliza datos de encuestas. Otro 23% (10 investigadores) utiliza los resultados de la medida de instrumentos. El 14% (6 investigadores) utiliza documentación diversa. El 7% (3 investigadores) utiliza imágenes digitales: fotografías; y diagramas. El 2% (1 investigador) utiliza audios digitales. Por último, otro 2% (1 investigador) utiliza historias clínicas.

Observamos que los docentes investigadores en UPC utilizan todos los tipos de datos de investigación descritos, principalmente de datos numéricos, por lo que requieren computadoras con software y memorias adecuadas.

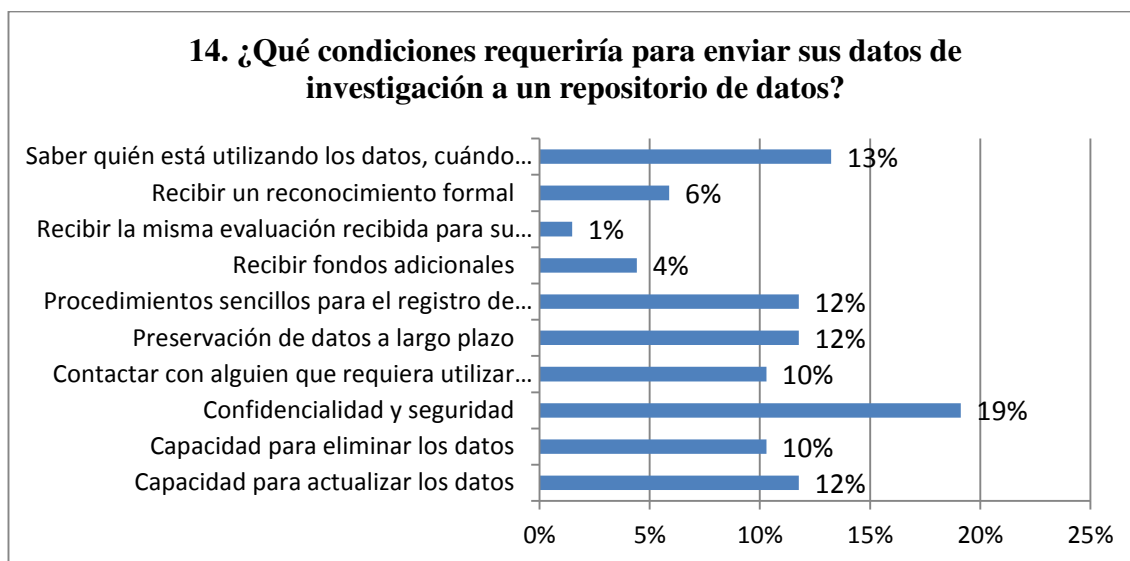
Bloque IV: almacenamiento de datos de investigación

Tabla 53 Distribución de frecuencias: condiciones requeridas por los investigadores de CS de la UPC para enviar sus datos de investigación a un repositorio

14. ¿Qué condiciones requeriría para enviar sus datos de investigación a un repositorio de datos?

Respuesta	fi	hi
Capacidad para actualizar los datos	8	12%
Capacidad para eliminar los datos	7	10%
Confidencialidad y seguridad	13	19%
Contactar con alguien que requiera utilizar mis datos	7	10%
Preservación de datos a largo plazo	8	12%
Procedimientos sencillos para el registro de los datos	8	12%
Recibir fondos adicionales	3	4%
Recibir la misma evaluación recibida para su publicación en una revista	1	1%
Recibir un reconocimiento formal	4	6%
Saber quién está utilizando los datos, cuándo y para qué fin	9	13%
Total	68	100%

Gráfico 54 Frecuencia relativa: condiciones requeridas por los investigadores de CS de la UNMSM para enviar sus datos de investigación a un repositorio



Se determinó que los investigadores requieren más de una condición para enviar sus datos de investigación a un repositorio de datos. El 19% (13 investigadores) requiere la condición de confidencialidad y seguridad. El 13% (9 investigadores) requiere la condición de saber quién está utilizando los datos, cuando y para qué fin. El 12% (8 investigadores) requiere la condición de capacidad para actualizar los datos. Otro 12% (8 investigadores) requiere la condición de preservación de datos a largo plazo. 12% más (8 investigadores) requiere la condición de que los procedimientos sean sencillos para el registro de los datos. Un 10% (7 investigadores) requiere la condición de que el repositorio tenga la capacidad para eliminar los datos. El 6% (4 investigadores) pondría sus datos de investigación en un repositorio si reciben un reconocimiento formal. Un 4% (3 investigadores) pondría sus datos de investigación a condición de recibir fondos adicionales. Y sólo el 1% (1 investigador) pondría como condición recibir la misma evaluación recibida para su publicación en una revista.

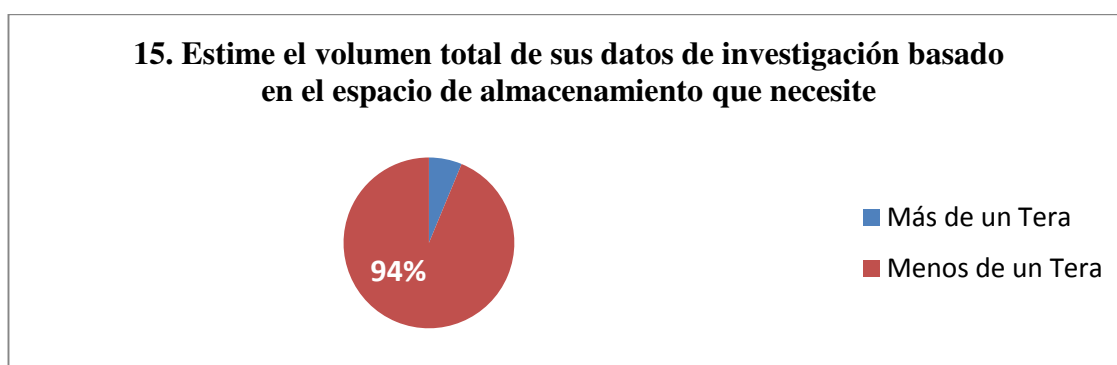
Es evidente, que existen muchas condiciones, lo importante es considerar y evaluar todas estas condiciones expuestas a fin de determinar la viabilidad de satisfacerlas.

Tabla 54 Distribución de frecuencias: espacio de almacenamiento que necesitan los investigadores de CS de la UPC

15. Estime el volumen total de sus datos de investigación basado en el espacio de almacenamiento que necesite (el promedio estimado por investigación).

Respuesta	fi	hi
Más de un Tera	1	6%
Menos de un Tera	15	94%
Total	16	100%

Gráfico 55 Frecuencia relativa: espacio de almacenamiento que necesitan los investigadores de CS de la UPC



Se determinó que el 94% (15 colaboradores) requieren como espacio de almacenamiento por investigación, menos de un Tera. Y sólo un mínimo 6% (un investigador) utiliza más de un Tera.

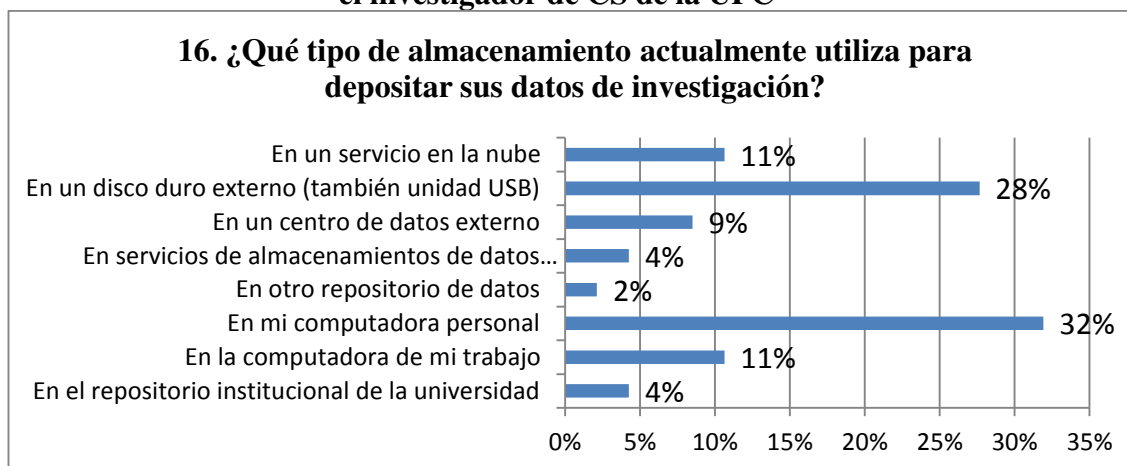
Esto se debe a que los investigadores cuando empiezan una investigación requieren liberar espacio, por lo general limpiando aplicaciones utilizadas, y guardando sus datos de investigación en otros dispositivos, como memorias externas o USB de gran capacidad. Por otro lado, muchos investigadores no utilizan archivos pesados como audios digitales (archivos sonoros, entrevistas etc.) o Imágenes digitales (fotografía, diagramas, etc.). Además, la tendencia para conservar archivos o datos de investigación es comprimir los mismos.

Tabla 55 Distribución de frecuencias: tipo de almacenamiento actualmente utiliza el investigador de CS de la UPC

16. ¿Qué tipo de almacenamiento actualmente utiliza para depositar sus datos de investigación?

Respuesta	fi	hi
En el repositorio institucional de la universidad	2	4%
En la computadora de mi trabajo	5	11%
En mi computadora personal	15	32%
En otro repositorio de datos	1	2%
En servicios de almacenamientos de datos como GITHUB	2	4%
En un centro de datos externo	4	9%
En un disco duro externo (también unidad USB)	13	28%
En un servicio en la nube	5	11%
Total	47	100%

Gráfico 56 Frecuencia relativa: tipo de almacenamiento que actualmente utiliza el investigador de CS de la UPC



Se determinó que los investigadores utilizan más de un tipo de almacenamiento. El 32% (15 investigadores) utiliza su computadora personal. El 28% utiliza un disco externo, así como, unidades USB. Un 11% (5 investigadores) deposita sus datos de investigación en la computadora de su trabajo. Otro 11% (5 investigadores) utiliza el servicio en la nube. Un 9% (4 investigadores) utiliza un centro de datos externos. Un 4% (2 investigadores) utilizan el repositorio institucional. Otro 4% (2 investigadores) utiliza el GITHUB. Y por último, un 2% (1 investigador) utiliza otro repositorio de datos.

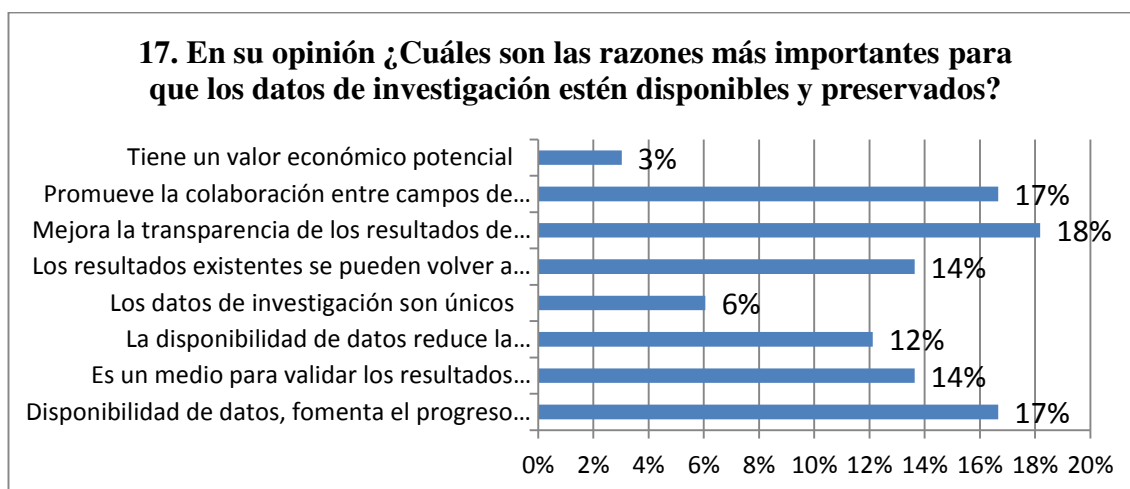
Como se puede observar los investigadores utilizan casi todos los tipos de almacenamiento o unidades de almacenamiento de datos disponibles, para leer y/o grabar sus datos. Por lo que la Universidad debe proveer estas herramientas, contando con un repositorio institucional que permita al investigador compartir, reutilizar y asegurar sus datos de investigación en una plataforma confiable.

Tabla 56 Distribución de frecuencias: razones más importantes para los investigadores de CS de la UPC para que sus datos de investigación estén disponibles y preservados

17. En su opinión ¿Cuáles son las razones más importantes para que los datos de investigación estén disponibles y preservados?

Respuesta	fi	hi
Disponibilidad de datos, fomenta el progreso de la ciencia (nueva investigación se basa en el conocimiento preexistente)	11	17%
Es un medio para validar los resultados obtenidos	9	14%
La disponibilidad de datos reduce la duplicación de esfuerzos de investigación	8	12%
Los datos de investigación son únicos	4	6%
Los resultados existentes se pueden volver a examinar.	9	14%
Mejora la transparencia de los resultados de la investigación.	12	18%
Promueve la colaboración entre campos de diferentes disciplinas.	11	17%
Tiene un valor económico potencial	2	3%
Total	66	100%

Gráfico 57 Frecuencia relativa: razones más importantes para los investigadores de CS de la UPC para que sus datos de investigación estén disponibles y preservados



Se determinó que los investigadores tienen más de una razón importante para que los datos de investigación estén disponibles y preservados. El 18% (12 investigadores) considera que la disponibilidad de datos mejora la transparencia de los resultados del investigador. El 17% (11 investigadores) entiende que una buena razón por la que sus datos de investigación podrían estar disponibles y preservados es que fomenta el progreso de la ciencia basado en el conocimiento preexistente. Otro 17% (11 investigadores) considera que la disponibilidad y preservación de los datos de investigación promueve la colaboración entre campos de diferentes disciplinas. Un 14% (9 investigadores) considera que es un medio para validar los resultados obtenidos. Otro 14% (9 investigadores) considera que al estar disponible los datos de investigación los resultados existentes se pueden volver a examinar. Un 12% (8 investigadores) entiende que la disponibilidad de datos reduce la duplicación de esfuerzos de investigación. Un 6% (4 investigadores) considera que los datos de investigación deben estar disponibles y preservados porque los datos de investigación son únicos. Y por último, sólo un 3% (2 investigadores) considera que tienen un valor económico potencial.

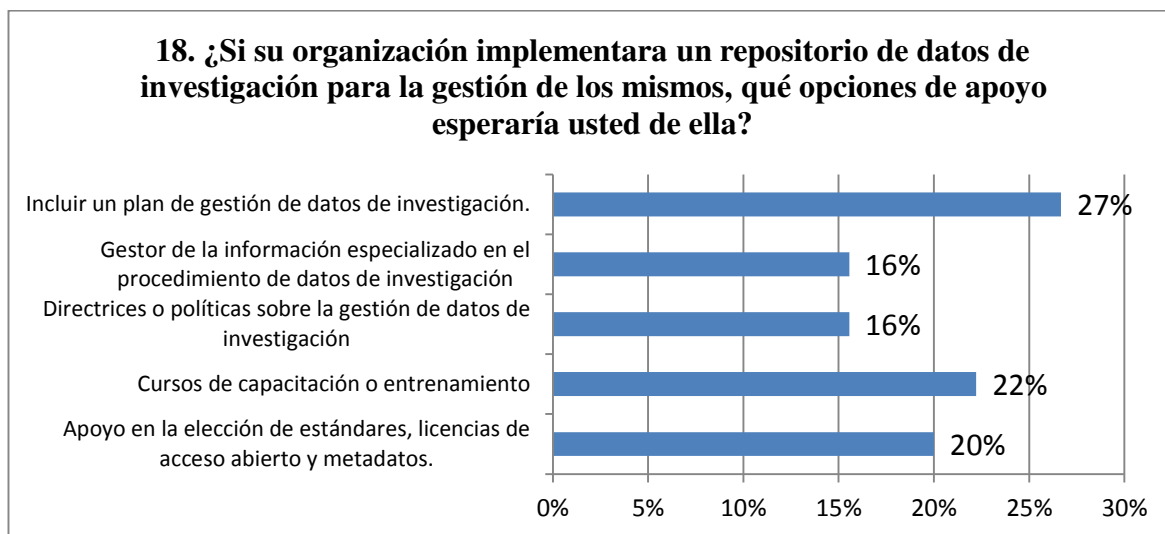
Debemos tomar en cuenta, que la disponibilidad y preservación de los datos asegura la calidad de los mismos, debido a su posibilidad de ser reutilizados por otros investigadores.

Tabla 57 Distribución de frecuencias: opciones de apoyo que el investigador de CS de la UPC esperaría de su institución si implementará un repositorio

18. ¿Si su organización implementara un repositorio de datos de investigación para la gestión de los mismos, qué opciones de apoyo esperaría usted de ella?

Respuesta	fi	hi
Apoyo en la elección de estándares, licencias de acceso abierto y metadatos.	9	20%
Cursos de capacitación o entrenamiento	10	22%
Directrices o políticas sobre la gestión de datos de investigación	7	16%
Gestor de la información especializado en el procedimiento de datos de investigación	7	16%
Incluir un plan de gestión de datos de investigación.	12	27%
Total	45	100%

Gráfico 58 Frecuencia relativa: opciones de apoyo que el investigador de CS de la UPC esperaría de su institución si implementará un repositorio



Se determinó que los investigadores esperan tener más de un tipo de apoyo de parte de su organización si se implementa un repositorio de datos de investigación. El 27% (12 investigadores) esperaría que en la implementación del repositorio se incluya un plan de gestión de datos de investigación. Un 22% (10 investigadores) esperaría apoyo a través de cursos de capacitación o entrenamiento. Un 20% (9 investigadores) esperaría apoyo en la elección de estándares, licencias de acceso abierto y meta datos, un 16% (7 investigadores) esperaría apoyo sobre directrices o políticas sobre la gestión de datos de investigación. Y otro 16% (7 investigadores) esperaría apoyo de un gestor de información especializado en el procedimiento de datos de investigación.

Podríamos decir que los apoyos naturales a los investigadores docentes son: espacio de trabajo; información bibliográfica especializada; evaluación y difusión de la producción científica; y apoyo multidisciplinario a los equipos de investigación. En ese sentido, todos los apoyos adicionales dirigidos a los investigadores docentes pueden tratarse como un reconocimiento hacia ellos.

Tabla 58 Distribución de frecuencias: preferencia del responsable para hacerse cargo del archivo de datos de investigación de los investigadores de CS de la UPC

19. ¿Quién le gustaría que sea el responsable del archivo de sus datos de investigación en un repositorio de datos?

Respuesta	fi	hi
Centro de TI	3	14%
Gestor de la Información	3	14%
Yo mismo	3	14%
Yo mismo, pero acompañado de un experto en datos de investigación (data curator)	13	59%
Total	22	100%

Gráfico 59 Frecuencia relativa: de la preferencia del responsable para hacerse cargo del archivo de datos de investigación de los investigadores de CS de la UPC



Se determinó que al 59% (13 investigadores) le gustaría ser, ellos mismos, los responsables de sus archivos, pero acompañado de un experto en datos de investigación (data curator). Un 14% (3 investigadores) le gustaría que el responsable de los archivos de sus datos de investigación sea un Centro de TI. Un 14% (3 investigadores) preferiría que sea un gestor de la información. Y por último un 14% (3 investigadores) le gustaría ser, el mismo, el responsable del archivo de sus datos de investigación.

Observamos que a la mayoría de investigadores les gustaría ser responsables de sus archivos de datos de investigación. Esta buena disposición de los investigadores por la disponibilidad y preservación de los datos de investigación, podría resultar una carga de trabajo adicional para el investigador, perjudicando su atención y

concentración sobre su investigación. En ese sentido, la gestión de datos de investigación, en un repositorio, es una función especializada que alcanza a la tecnología, métodos, procesos y procedimientos pertinentes para la recolección, selección, registro, preservación, disponibilidad y reutilización de los datos de investigación.

5.3. Análisis Comparativo e Interpretación de UNMSM – UPC

Bloque I: información general

Tabla 59 Distribución de frecuencias: instituciones donde los investigadores de la UNMSM - UPC realizan sus trabajos de investigación

1. Seleccione la institución en la que realiza trabajos de investigación.

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
Científica del Sur	1	5%	0	0%	1	3%
Hospital Nacional 2 de mayo	1	5%	0	0%	1	3%
Instituto Nacional Cardiovascular	0	0%	1	8%	1	3%
Instituto Nacional de Salud	0	0%	1	8%	1	3%
Ricardo Palma	1	5%	0	0%	1	3%
Superintendencia Nacional de Salud	1	5%	0	0%	1	3%
TU Kaiserslautern (Alemania)	1	5%	0	0%	1	3%
UNMSM - Universidad Nacional Mayor de San Marcos	0	0%	9	75%	9	27%
UPC - Universidad Peruana de Ciencias Aplicadas	15	71%	1	8%	16	48%
UPCH - Universidad Peruana Cayetano Heredia	1	5%	0	0%	1	3%
TOTAL DE RESPUESTAS	21	100%	12	100%	33	100%

Se determinó que hay investigadores que realizan trabajos de investigación en más de una institución. Se determinó que 10 instituciones financian investigaciones en el campo de las Ciencias de la Salud, a los 25 docentes investigadores encuestados. Se determinó que hay 7 instituciones que financian 21 investigaciones, 16 de las cuales se realizan en la UPC, a un promedio de 3 investigaciones por institución, mientras que 4 instituciones financian 12 investigaciones, 9 de las cuales se realizan en la UNMSM, también a un promedio de 3 investigaciones por institución. Se determinó que el 75% (9 investigadores) realizan sus trabajos de investigación en la misma UNMSM, y un 25% (3 investigadores) lo hace en instituciones privadas, mientras que

los docentes investigadores de la UPC en un 71% (15 investigadores) lo hace en la misma universidad. Se determinó que del 48% (16 investigadores) del total de 25 investigadores encuestados que realiza sus investigaciones en la UPC, el 8% (1 investigador) también lo hace en la UNMSM.

La preferencia de los investigadores por realizar sus investigaciones en la misma institución se debe a su calidad de docentes, es decir, investigan en su centro de trabajo, esta circunstancia evita que el investigador se traslade de un lugar a otro, evitando la pérdida de un recurso valioso como es el tiempo, del mismo modo, la condición de docente le permite al investigador estar en contacto permanente con el mundo académico y a su vez poder utilizar los recursos tecnológicos, científicos y financieros que la Universidad consigue a través de convenios y acuerdos con instituciones u organizaciones nacionales e internacionales.

Tabla 60 Distribución de frecuencias: líneas de investigación a las que se dedican los investigadores de CS de la UNMSM - UPC

2. ¿Cuáles son las líneas de investigación a las que usted se dedica?

Respuesta	UPC		UNMSM		Ambos	
	Fi	hi	fi	hi	fi	hi
Biomateriales	1	3%	0	0%	1	2%
Bioquímica, Biología Molecular y Celular	4	13%	1	10%	5	12%
Cirugía plástica	1	3%	0	0%	1	2%
Clinica	1	3%	0	0%	1	2%
Educación en Ciencias de la Salud	4	13%	3	30%	7	17%
Enfermedades Infecciosas	0	0%	1	10%	1	2%
Epidemiología Clínica y Salud Basada en Evidencias	6	19%	1	10%	7	17%
Epidemiología e Informática en Salud	0	0%	1	10%	1	2%
Medicina Clínica	1	3%	0	0%	1	2%
Nutrición, Deporte y Actividad Física	2	6%	1	10%	3	7%
Personalidad y Desarrollo	0	0%	1	10%	1	2%
Psicología deportiva	1	3%	0	0%	1	2%
Psicología educativa	1	3%	0	0%	1	2%
Psicología social y de las organizaciones	2	6%	0	0%	2	5%
Salud mental	1	3%	0	0%	1	2%
Salud Pública y Gestión en Salud	4	13%	1	10%	5	12%
Sistemas Computacionales Embebidos	1	3%	0	0%	1	2%
Tratamientos de Terapia Física	1	3%	0	0%	1	2%
TOTAL DE RESPUESTAS	31	100%	10	100%	41	100%

De acuerdo al planteamiento del ciclo de vida de los datos de investigación las líneas de investigación están asociadas a la etapa de creación o generación de datos.

Se determinó que hay investigadores que tienen más de una línea de investigación. Se determinó que hay 18 líneas de investigación en el campo de las Ciencias de la Salud y 41 investigaciones en marcha (31 en la UPC y 10 en la UNMSM). Se determinó que 16 investigadores de la UPC realizan 31 investigaciones a un promedio de 1.94 investigaciones por investigador, mientras que 9 investigadores de la UNMSM realizan 10 investigaciones a un promedio de 1.11 investigaciones por investigador. Esto se explica, en la UPC, porque la investigación es un lineamiento estratégico, lo que ha inducido, en los últimos años, ha dirigir más recursos hacia la investigación. Se determinó que el 17% (7 investigadores) de un total de 25 investigadores, tiene como línea de investigación la Educación en Ciencias de la Salud. Otro 17% (7 investigadores) de un total de 25 investigadores tiene como línea de investigación la Epidemiología Clínica y Salud basada en Evidencias. Se determinó que hay 12 investigadores que desarrollan sólo una línea de investigación.

Tabla 61 Distribución de frecuencias: agencias que contribuyen con fondos a la investigación en UNMSM - UPC

3. ¿Cuál de las siguientes opciones describe mejor la principal agencia de fondos para su investigación?

Respuesta	UPC		UNMSM		Ambos	
	Fi	hi	fi	Hi	fi	hi
Entidad privada	5	26%	2	20%	7	24%
Gobierno nacional	4	21%	2	20%	6	21%
Mi institución	10	53%	6	60%	16	55%
TOTAL DE RESPUESTAS	19	100%	10	100%	29	100%

Se determinó que hay investigadores que perciben que hay más de una agencia con fondos para su investigación. El 55% (16 investigadores) describe como la principal agencia de fondos para su investigación a su institución. El 24% (7 investigadores) considera que la principal agencia de fondos para su investigación son las entidades privadas. Y un 21% (6 investigadores) considera que es el gobierno nacional es la principal agencia de fondos para su investigación.

Debemos resaltar, la participación casi a la par de la entidad privada con el gobierno nacional, esto evidencia que hay una gran preocupación de parte del gobierno nacional por investigar en un campo como es el de las ciencias de la salud, ya que para cualquier gobierno la salud de la población o de su comunidad debe ser prioritaria, en

ese sentido, debería de alguna manera el gobierno nacional apoyar a las entidades privadas y también a las universidades nacionales o privadas para que la investigación en este campo se potencie.

Bloque II: uso de datos de investigación

Tabla 62 Distribución de frecuencias: uso que le da el investigador de la UNMSM - UPC

4. ¿Qué hace usted con sus datos de investigación?

Respuesta	UPC		UNMSM		Ambos	
	Fi	hi	fi	Hi	fi	hi
Los conserva	16	100%	9	100%	25	100%
Los desecha	0	0%	0	0%	0	0%
Total	16	100%	9	100%	25	100%

Se determinó que el 100% (25 investigadores) tanto de la UNMSM como de la UPC, conserva sus datos de investigación.

Tabla 63 Distribución de frecuencias: tiempo que conserva el investigador de CS de la UNMSM - UPC sus datos de investigación

4.a En caso su respuesta sea "Los conserva", ¿Cuánto tiempo los conservaría?

Respuesta	UPC		UNMSM		Ambos	
	Fi	hi	fi	Hi	fi	hi
Más de 1 año	16	100%	7	78%	23	92%
Menos de 1 año	0	0%	2	22%	2	8%
Total	16	100%	9	100%	25	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación el uso que le da el investigador a sus de datos de investigación y el tiempo que conserva sus datos de investigación están asociados a las etapas de generación y preservación de datos.

Para asegurar que las estrategias básicas de difusión despierten el interés, en los docentes investigadores, en compartir y reutilizar los datos de investigación se

requiere puntualizar y repetir sistemáticamente los beneficios del acceso abierto a la información científica, como parte de este aseguramiento se requiere de realizar actividades de: medición sistemática de las respuestas a los mensajes; seguimiento de los procesos de inducción y capacitación; comparación de estándares de las redes sociales.

Tabla 64 Distribución de frecuencias: cantidad de datos de investigación que el investigador de CS de la UNMSM - UPC coloca disponibles a los demás

5. ¿Cuánto de sus datos de investigación los coloca disponibles a los demás?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	Hi	fi	hi
Algunos	8	50 %	8	89%	16	64%
Ninguno	6	38%	1	11%	7	28%
Todos	2	13%	0	0%	2	8%
Total	16	100%	9	100%	25	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación la cantidad de datos de investigación que el investigador coloca disponibles a los demás está asociado a la etapa de acceso a datos de investigación.

Se determinó que mientras el 50% (8 investigadores) de la UPC coloca algunos de sus datos disponibles a los demás, los investigadores de la UNMSM, lo hace en un 89% (8 investigadores). El 13% (2 investigadores) de la UPC coloca todos sus datos de investigación a disposición de los demás. El 64% (16 investigadores) de ambas universidades, coloca algunos de sus datos de investigación disponibles a los demás. El 28% (7 investigadores) no coloca ningún dato de investigación. Y el 8% (2 investigadores) colocan todos sus datos de investigación a disposición de los demás.

De lo observado se desprende, que los investigadores están dispuestos a compartir sus datos investigación con los demás, pero no lo hacen debido a que todavía no hay un medio o una plataforma para hacerlo con confianza en el cual puedan intercambiar los datos.

Tabla 65 Distribución de frecuencias: usuarios que tienen acceso a los datos de investigación de los investigadores de CS de la UNMSM - UPC

6. ¿A quién concede usted el acceso a sus datos de investigación?

Respuesta	UPC		UNMSM		Ambos	
	Fi	hi	fi	Hi	fi	hi
A la comunidad científica	4	15%	3	21%	7	18%
A las personas interesadas en solicitar mis datos	11	42%	8	57%	19	48%
A los investigadores de mi misma disciplina	6	23%	3	21%	9	23%
A todos los miembros de mi institución	4	15%	0	0%	4	10%
Al público	1	4%	0	0%	1	3%
Total	26	100%	14	100%	40	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación los usuarios que tienen acceso a los datos de investigación de los investigadores está asociado a la etapa de acceso a datos de investigación.

Se determinó que el 48% (19 investigadores) del total de 25 investigadores encuestados, concede acceso a sus datos de investigación a las personas que se lo solicitan. El 23% (9 investigadores) lo hace a los investigadores de su misma disciplina. El 18% (7 investigadores) lo hace a la comunidad científica. El 10% (4 investigadores) lo hace a todos los miembros de su institución. Y por último el 3% (1 investigador) lo hace al público en general.

Se observa que la mayoría de los investigadores están dispuestos a conceder acceso a sus datos de información si se lo solicitan, es evidente que de parte de ellos hay un gran sentido de propiedad y protección sobre, algo que es importante tener en cuenta, los derechos de autor, la propiedad intelectual, la privacidad y la seguridad.

Tabla 66 Distribución de frecuencias: medios a través de los cuales se accede a los datos de investigación de los investigadores de CS de la UNMSM - UPC

7. ¿A través de que medio pueden acceder otros investigadores a los datos de su investigación?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
Como material complementario vinculado a sus publicaciones	5	19%	3	20%	8	20%
Dispositivos electrónicos (USB, CD-Rom, etc.) y/o correo electrónico	9	35%	3	20%	12	29%
Repositorio de datos	6	23%	4	27%	10	24%
Web personal o institucional	6	23%	5	33%	11	27%
Total	26	100%	15	100%	41	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación los medios a través de los cuales se accede a los datos de investigación están asociados a la etapa de acceso a datos de investigación.

Se determinó que los investigadores utilizan más de un medio para que otros investigadores puedan acceder a sus datos de investigación. Se determinó que el 29% (12 investigadores) del total de 25 investigadores encuestados, dan acceso a otros investigadores a sus datos de investigación a través de dispositivos electrónicos (USB, CD-Rom, etc.) y/o correo electrónico. El 27% (11 investigadores) utiliza una Web personal o institucional. El 24% (10 investigadores) utiliza un repositorio de datos. Y el 20% (8 investigadores) utiliza sus datos de investigación como material complementario vinculado a sus publicaciones, para facilitar el acceso a sus datos de investigación a otros investigadores.

En la actualidad podemos encontrar una gran oferta de dispositivos electrónicos y medios informáticos, que permiten poder intercambiar o compartir información de manera fácil, rápida y fluida. En ese sentido, los investigadores que de manera permanente generan datos de investigación tienen la opción de elegir el más conveniente.

Tabla 67 Distribución de frecuencias: utilización de datos generados por otros investigadores o instituciones, por los investigadores de CS de la UNMSM - UPC

8. ¿Utiliza los datos generados por otros investigadores o instituciones cómo OMS u OPS?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
No	7	44%	2	22%	9	36%
Si	9	56%	7	78%	16	64%
Total	16	100 %	9	100%	25	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación la utilización de datos generados por otros investigadores o instituciones y los medios a través de los cuales se accede a los datos de investigación están asociados a la etapa de acceso a datos de investigación, a la etapa reutilización y a la etapa de difusión.

Se determinó que el 64% (16 investigadores) de los 25 investigadores encuestados sí utiliza los datos generados por otros investigadores o instituciones como OMS u OPS. Y el 36% (9 investigadores) no lo hace.

El Internet contribuye o facilita a los investigadores acceder a catálogos y bases de datos especializados, donde se puede encontrar investigaciones a texto parcial o completo en la misma línea de investigación, como es el caso del uso de los datos generados en instituciones como la OMS y OPS. Sin embargo, en el caso de los investigadores que no utilizan los datos generados por otros investigadores se debe a la falta de información sobre la posibilidad de utilizar los datos generados por otros investigadores sin contraer ningún tipo de compromiso o a las dificultades que tienen para acceder a los datos de investigación publicados.

Tabla 68 Distribución de frecuencias: reutilización de datos generados por otros investigadores o instituciones, por los investigadores de CS de la UNMSM - UPC

9. ¿Reutiliza usted los datos de otros investigadores?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
No	11	69%	7	78%	18	72%
Si	5	31%	2	22%	7	28%
Total	16	100%	9	100%	25	100%

Se determinó que el 28% (7 investigadores) sí reutiliza los datos de investigación de otros investigadores. Y el 72% (18 investigadores) no reutiliza los datos de investigación de otros investigadores.

En nuestro país todavía hay dificultades para reutilizar los datos de investigación de otros investigadores, esto se debe, principalmente, a que no todos los portales, webs o repositorios están diseñados para poner a disposición de manera idónea los datos de investigación que conservan, también se debe a políticas restrictivas, otorgamiento de licencias o limitaciones técnicas. La reutilización de datos de investigación se refiere a las facilidades que se dan para que un investigador pueda usar las veces que desee o que crea conveniente un dato de investigación, que se encuentra en un repositorio o en una base de datos de acceso abierto, la finalidad de la reutilización es promover y facilitar la investigación y consecuentemente la innovación, racionalizando el uso de recursos escasos. En este caso, los investigadores no reutilizan los datos de investigación de otros investigadores, porque desconocen su existencia o no tienen los medios tecnológicos para hacerlo

Tabla 69 Distribución de frecuencias: tipo de licencia que aplican los investigadores de CS de la UNMSM - UPC

9.a. En caso su respuesta sea afirmativa, seleccione el tipo de licencia que aplicaría.

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
Acuerdos de cooperación	1	14%	0	0%	1	11%
Creative Commons	1	14%	2	100%	3	33%
Ninguno	1	14%	0	0%	1	11%
No sé	2	29%	0	0%	2	22%
Open Data License Attribution	2	29%	0	0%	2	22%
Total	7	100%	2	100%	9	100%

La obtención de licencias establecidas para los datos abiertos, está asociada a la etapa de reutilización y de difusión del ciclo de vida de los datos.

Se determinó que más de un investigador aplicaría a un tipo de licencia. El 33% (3 investigadores) aplicarían a la licencia Creative Commons. Un 22% (2 investigadores) aplicaría a Open Data License Attribution. Otro 22% (2 investigadores) no sabe a que licencia aplicar. El 11% (1 investigador) aplicaría a licencias fundadas en acuerdos de cooperación. Y otro 11% (1 investigador) no aplicaría a ninguna licencia.

Un aspecto a tomar en cuenta es que las licencias, que se otorgan para el uso de datos de investigación o de información especializada, están propuestas para defender los derechos de autor y por lo tanto la propiedad intelectual. Otro aspecto importante a considerar es que los investigadores son creadores de conocimiento, a partir de un trabajo científico y académico.

Bloque III: organización de datos de investigación

Tabla 70 Distribución de frecuencias: sobre como registra sus datos de investigación los investigadores de CS de la UNMSM - UPC

10. ¿Cómo registra usted los datos de su investigación?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
No utilizo	3	14%	1	10%	4	13%
Utilizando alguna nomenclatura	4	19%	1	10%	5	16%
Utilizando estándares internacionales	5	24%	0	0%	5	16%
Utilizando un software adecuado	9	43%	8	80%	17	55%
Total	21	100%	10	100%	31	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación el registro de los datos está relacionado al análisis e interpretación y asociado a las etapas de evaluación y reutilización de los datos.

Se determinó que los investigadores utilizan más de un medio para registrar sus datos de investigación. El 55% (17 investigadores) de los 25 investigadores encuestados utilizarían un software adecuado para registrar sus datos. Un 16% (5 investigadores) registraría sus datos utilizando alguna nomenclatura. Otro 16% (5 investigadores) utiliza estándares internacionales. Y por último un 13% (4 investigadores) no utiliza ningún registro para sus datos de investigación.

Sí conservar los datos de investigación es importante, más importante aún es como se registran. Es evidente que para que los investigadores registren adecuadamente sus datos de investigación, deben necesariamente conocer los procedimientos o protocolos para este fin.

Tabla 71 Distribución de frecuencias: tipo de formato que usa el investigador de CS de la UNMSM - UPC para almacenar sus datos de investigación

11. ¿Qué tipo de formato usa cuando almacena datos de investigación?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
Aplicaciones de software	2	5%	0	0%	2	3%
Audio (MP3, WAV, AIFF, OGG, etc.)	1	2%	1	5%	2	3%
Bases de datos	1	2%	0	0%	1	2%
Bases de datos (MS Access, MySQL, Oracle, NON-SQL, etc.)	3	7%	0	0%	3	5%
Datos de configuración (INI, CONF etc.)	1	2%	0	0%	1	2%
Documentos de texto (DOC, ODF, TXT, etc.)	7	16%	7	32%	14	22%
Documentos Digitales (PDF, LATEX, etc.)	6	14%	4	18%	10	15%
DTA	2	5%	0	0%	2	3%
Gráficos / Imágenes (JPEG, SVG, PNG, GIF, TIFF, etc.)	6	14%	2	9%	8	12%
Hojas de cálculo (XLS, ODS, CSV, SAS, etc.)	8	19%	7	32%	15	23%
Software de código de las aplicaciones de código (CSS, JavaScript, Java, etc.)	1	2%	0	0%	1	2%
Stata	2	5%	0	0%	2	3%
Texto estructurado (HTML, JSON, TEX, XML, etc.)	0	0%	1	5%	1	2%
Videos / películas (MPEG, AVI, WMV, MP4, etc.)	3	7%	0	0%	3	5%
Total	43	100%	22	100%	65	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación el tipo de formato que usa el investigador para almacenar sus datos tiene que ser adecuado para ser reutilizado.

Se determinó que los investigadores usan más de un formato para almacenar sus datos de investigación. Se determinó que 16 investigadores de la UPC, utilizan en total 43 formatos para almacenar sus datos de investigación, con un promedio de 2,68 formatos por investigador, mientras que 9 investigadores de la UNMSM utilizan 22 formatos para almacenar sus datos de investigación, con un promedio de 2,44 por investigador. El 23% (15 investigadores) del total de 25 investigadores encuestados utiliza hojas de cálculo como formato para almacenar sus datos de investigación. El 22% (14 investigadores) utiliza documentos de texto. El 15% (10 investigadores) utiliza documentos digitales. El 12% (8 investigadores) utiliza gráficos e imágenes. El

5% (3 investigadores) utiliza bases de datos. Otro 5% (3 investigadores) utiliza vídeos y películas. El 3% (2 investigadores) utiliza aplicaciones de software. El 3% (2 investigadores) utiliza audios. El 3% (2 investigadores) utiliza DTA. Otro 3% (2 investigadores) utiliza Stata. Y por último un 2% (1 investigador) utiliza texto estructurado.

En general, los formatos se eligen de acuerdo al contenido. El investigador debe tener cierto conocimiento sobre cómo elegir un estándar para almacenar cada uno de sus datos de investigación teniendo en cuenta el uso de formatos adecuados para la reutilización de los mismos.

Tabla 72 Distribución de frecuencias: técnicas de recolección de datos que utiliza el investigador de CS de la UNMSM - UPC

12. ¿Qué técnicas de recolección de datos utiliza?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
Encuesta	10	33%	6	30%	16	32%
Entrevista	4	13%	5	25%	9	18%
Experimental (implica un cierto grado de manipulación)	8	27%	3	15%	11	22%
Historias clínicas	1	3%	0	0%	1	2%
Observacional (no incluida la manipulación)	7	23%	6	30%	13	26%
Total	30	100%	20	100%	50	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación las técnicas de recolección de datos que utiliza el investigador están asociadas a la etapa de creación y almacenamiento.

Se determinó que los investigadores utilizan más de una técnica de recolección de datos. Los investigadores de la UPC utilizan las 5 técnicas propuestas mientras que los investigadores de la UNMSM utilizan sólo 4, no utilizan historias clínicas en sus investigaciones. Se determinó que los 16 investigadores de la UPC utilizaron en total 30 veces técnicas de recolección, con un promedio de 1,87 técnicas utilizadas por investigador mientras que 9 investigadores de la UNMSM utilizaron en total 20 veces técnicas de recolección con un promedio de 2,22 técnicas utilizadas por investigador.

Se determinó que el 32% (16 investigadores) utilizó la encuesta como técnica para la recolección de datos. El 22% (11 investigadores) utilizó la técnica experimental. El 18% (9 investigadores) utilizó la entrevista. Y por último el 2% (un investigador) utiliza como técnica las historias clínicas.

Lo cierto es que existen varias técnicas de recolección de datos, principalmente las que obtienen información de fuentes primarias, lo importante, antes de elegir una técnica de recolección de datos, es determinar la naturaleza de la unidad de estudio, el tamaño de la muestra y los recursos para llevar a cabo dicha recolección. Sin embargo, en este estudio se determinó que los investigadores en Ciencias de la Salud de ambas universidades utilizan la técnica de la encuesta, dado que les permite conocer una determinada situación a partir de fuentes primarias.

Tabla 73 Distribución de frecuencias: tipo de datos de investigación que utilizan los investigadores de CS de la UNMSM - UPC

13. ¿Qué tipo de datos de investigación utiliza usted en sus investigaciones?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
Audios digitales (archivos sonoros, entrevistas)	1	2%	1	4%	2	3%
Datos de encuestas	10	23%	7	30%	17	26%
Datos numéricos	12	28%	6	26%	18	27%
Documentación diversa	6	14%	4	17%	10	15%
Historias clínicas	1	2%	0	0%	1	2%
Imágenes digitales (fotografía, diagramas)	3	7%	0	0%	3	5%
Resultados de la medida de instrumentos	10	23%	5	22%	15	23%
Total	43	100%	23	100%	66	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación el tipo de datos de investigación, requiere de análisis e interpretación y por lo tanto está asociado a la etapa de evaluación.

Se determinó que los investigadores utilizan más de un tipo de datos de investigación. Se determinó que los 16 investigadores de la UPC utilizaron en total 43 tipo de datos, con un promedio de 2,68 tipos de datos por investigador mientras que 9

investigadores de la UNMSM utilizaron en total 23 tipos de datos, con un promedio de 2,55 tipos de datos por investigador. Se determinó que el 27% (18 investigadores) utiliza datos numéricos. El 26% (17 investigadores) utiliza datos de encuestas. El 23% (15 investigadores) utiliza resultados de la medida de instrumentos. El 15% (10 investigadores) utiliza documentación diversa. El 5% (3 investigadores) utiliza imágenes digitales como fotografías y diagramas. Y el 2% (1 investigador) utiliza historias clínicas.

Se observa que los investigadores utilizan más de un tipo de datos de investigación, debido a qué durante el desarrollo de una investigación, sobre todo de investigaciones experimentales, se va requiriendo los diferentes tipos de datos de investigación.

Bloque IV: almacenamiento de datos de investigación

Tabla 74 Distribución de frecuencias: condiciones requeridas por los investigadores de CS de la UNMSM - UPC para enviar sus datos de investigación a un repositorio

14. ¿Qué condiciones requeriría para enviar sus datos de investigación a un repositorio de datos?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
Capacidad para actualizar los datos	8	12%	5	14%	13	12%
Capacidad para eliminar los datos	7	10%	4	11%	11	10%
Confidencialidad y seguridad	13	19%	6	16%	19	18%
Contactar con alguien que requiera utilizar mis datos	7	10%	2	5%	9	9%
Preservación de datos a largo plazo	8	12%	2	5%	10	10%
Procedimientos sencillos para el registro de los datos	8	12%	5	14%	13	12%
Recibir fondos adicionales	3	4%	3	8%	6	6%
Recibir la misma evaluación recibida para su publicación en una revista	1	1%	2	5%	3	3%
Recibir un reconocimiento formal	4	6%	3	8%	7	7%
Saber quién está utilizando los datos, cuándo y para qué fin	9	13%	5	14%	14	13%
Total	68	100%	37	100%	105	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación las condiciones que exigen los investigadores para compartir sus datos de investigación están en las etapas de almacenamiento y de difusión.

Se determinó que los investigadores requieren más de una condición para enviar sus datos de investigación a un repositorio de datos. Las condiciones expuestas son 10. El total de condiciones solicitadas por los 25 investigadores son 105 (las condiciones se repiten en promedio 10.5 veces). Los 16 investigadores de la UPC en total solicitan 68 condiciones, a un promedio de 4,25 condiciones por investigador, mientras que los 9 investigadores de la UNMSM en total solicitan 37 condiciones, a un promedio de 4,11 por investigador. Se determinó que el 18% (19 investigadores) requiere como condición, confidencialidad y seguridad. El 13% (14 investigadores) requiere como condición, saber quién está utilizando los datos, cuando y para qué fin. El 12% (13 investigadores) requiere como condición, capacidad para actualizar sus datos. El 10% (11 investigadores) requiere como condición, la capacidad para eliminar los datos. Otro 10% (11 investigadores) requiere como condición, la preservación de datos a largo plazo. El 9% (9 investigadores) requiere como condición, contactar con alguien que requiera utilizar sus datos. El 7% (7 investigadores) requiere como condición recibir un reconocimiento formal. El 6% (6 investigadores) requiere como condición recibir fondos adicionales. Y sólo el 3% requiere como condición recibir la misma evaluación recibida para su publicación en una revista.

Los repositorios u otros medios especializados en la gestión de datos de investigación, sobre todo los de acceso abierto, son de suma utilidad, una característica sustantiva es que permiten centralizar datos de investigación, de acuerdo a la especialización línea de investigación, que pueden compartirse o reutilizarse para mejorar una investigación existente o generar una nueva investigación.

Tabla 75 Distribución de frecuencias: espacio de almacenamiento que necesita los investigadores de CS de la UNMSM - UPC

15. Estime el volumen total de sus datos de investigación basado en el espacio de almacenamiento que necesite (el promedio estimado por investigación).

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
Más de un Tera	1	6%	0	0%	1	4%
Menos de un Tera	15	94%	9	100%	24	96%
Total	16	100%	9	100%	25	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación el volumen total de datos de investigación está asociado a la etapa de almacenamiento.

Se determinó que el 96% (24 investigadores) utiliza menos de un Tera como espacio de almacenamiento para guardar sus datos de investigación. Y sólo un 4% (1 investigador) utiliza más de un Tera.

Los investigadores deben gestionar el espacio de almacenamiento para sus datos de investigación, principalmente en sus computadoras personales y por lo tanto tienen que proyectar el volumen de datos de investigación que manejarán durante el desarrollo de su investigación, está claro, que hay una relación directa entre el número de investigaciones que realiza y el volumen de almacenamiento que necesita, así como el tamaño de su investigación con el volumen de almacenamiento.

Tabla 76 Distribución de frecuencias: tipo de almacenamiento actualmente utiliza el investigador de CS de la UNMSM - UPC

16. ¿Qué tipo de almacenamiento actualmente utiliza para depositar sus datos de investigación?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
En el repositorio institucional de la universidad	2	4%	2	9%	4	6%
En la computadora de mi trabajo	5	11%	3	14%	8	12%
En mi computadora personal	15	32%	9	41%	24	35%
En otro repositorio de datos	1	2%	0	0%	1	1%
En servicios de almacenamientos de datos como GITHUB	2	4%	0	0%	2	3%
En un centro de datos externo	4	9%	1	5%	5	7%
En un disco duro externo (también unidad USB)	13	28%	4	18%	17	25%
En un servicio en la nube	5	11%	3	14%	8	12%
Total	47	100%	22	100%	69	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación las condiciones de almacenamiento están asociadas a la etapa de preservación.

Se determinó que los investigadores utilizan más de un tipo de almacenamiento para depositar sus datos de investigación. Los tipos de almacenamiento expuestos son 8. El total de tipos de almacenamiento utilizados por los 25 investigadores son 69 (los tipos de almacenamiento se repiten en promedio 2,76 veces). Los 16 investigadores de la UPC en total utilizan 47 tipos de almacenamiento, a un promedio de 2,94 tipo de almacenamiento por investigador, mientras que los 9 investigadores de la UNMSM en total utilizan 22 tipos de almacenamiento, a un promedio de 2,44. Se determinó que el 35% (24 investigadores) utiliza para almacenar sus datos su computadora personal. El 25% (17 investigadores) lo hace en un disco duro externo o un USB. El 12% (8 investigadores) lo hace en la computadora de su trabajo. Otro 12% (8 investigadores) utiliza los servicios en la nube. El 7% (5 investigadores) utiliza un centro de datos externo. El 6% (4 investigadores) utiliza el repositorio institucional de la Universidad. El 3% (2 investigadores) utiliza los servicios de almacenamientos de datos como GITHUB. Y por último 1% (1 investigador) utiliza otro repositorio de datos.

El tipo de almacenamiento o dispositivo de almacenamiento depende del tipo de dato de investigación que se va a almacenar y de que los datos de investigación se almacenen de manera temporal o permanente. En este sentido es necesario contar con un programa de difusión y comunicación sobre los beneficios de presentar los datos en un repositorio de datos.

Tabla 77 Distribución de frecuencias: razones más importantes para los investigadores de CS de la UNMSM - UPC para que sus datos de investigación estén disponibles y preservados

17. En su opinión ¿Cuáles son las razones más importantes para que los datos de investigación estén disponibles y preservados?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
Disponibilidad de datos, fomenta el progreso de la ciencia (nueva investigación se basa en el conocimiento preexistente)	11	17%	6	18%	17	17%
Es un medio para validar los resultados obtenidos	9	14%	5	15%	14	14%
La disponibilidad de datos reduce la duplicación de esfuerzos de investigación	8	12%	8	24%	16	16%
Los datos de investigación son únicos	4	6%	0	0%	4	4%
Los resultados existentes se pueden volver a examinar.	9	14%	2	6%	11	11%
Mejora la transparencia de los resultados de la investigación.	12	18%	8	24%	20	20%
Promueve la colaboración entre campos de diferentes disciplinas.	11	17%	4	12%	15	15%
Tiene un valor económico potencial	2	3%	0	0%	2	2%
Total	66	100%	33	100%	99	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación, la disponibilidad de los datos está relacionada con la etapa de transformación y ésta necesariamente está asociada a la etapa de preservación.

Se determinó que los investigadores tienen más de una razón para que sus datos de investigación estén disponibles y preservados. Las razones expuestas son 8. El total de razones utilizadas por los 25 investigadores son 99 (las razones se repiten en promedio 3,96 veces). Los 16 investigadores de la UPC en total utilizan 66 razones a un promedio de 4,13 razones por investigador, mientras que los 9 investigadores de la UNMSM en total utilizan 33 razones, a un promedio de 3,66 razones por investigador. Se determinó que el 20% (20 investigadores) considera como una razón importante,

para que sus datos de investigación estén disponibles y preservados, la mejora en la transparencia de los resultados de la investigación. El 17% (17 investigadores) considera a la disponibilidad de datos como una razón importante porque fomenta el progreso de la ciencia. El 16% (16 investigadores) considera a la disponibilidad de datos como una razón importante porque reduce la duplicación de esfuerzos de investigación. El 15% (15 investigadores) considera a la disponibilidad de datos como una razón importante porque permite la colaboración entre campos de diferentes disciplinas. El 14% (14 investigadores) considera a la disponibilidad de datos como una razón importante porque considera que es un medio para validar los resultados obtenidos. El 11% (11 investigadores) considera a la disponibilidad de datos como una razón importante porque los resultados existentes se pueden volver a examinar. Y por último, el 2% (2 investigadores) considera a la disponibilidad de datos como una razón importante porque tiene un valor económico potencial.

La disponibilidad de datos de investigación además de evitar un doble esfuerzo en su búsqueda y promover la transparencia, fomenta y contribuye a la investigación, lo que tiene como consecuencia la generación de nuevo conocimiento. Un conocimiento que tendría que estar disponible y preservado y a su vez constituirse en una cadena sin fin de generación, producción y desarrollo de conocimientos.

Tabla 78 Distribución de frecuencias: opciones de apoyo que el investigador de CS de la UNMSM - UPC esperaría de su institución si implementará un repositorio

18. ¿Si su organización implementara un repositorio de datos de investigación para la gestión de los mismos, qué opciones de apoyo esperaría usted de ella?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
Apoyo en la elección de estándares, licencias de acceso abierto y metadatos.	9	20%	5	16%	14	18%
Cursos de capacitación o entrenamiento	10	22%	7	22%	17	22%
Directrices o políticas sobre la gestión de datos de investigación	7	16%	5	16%	12	16%
Gestor de la información especializado en el procedimiento de datos de investigación	7	16%	8	25%	15	19%
Incluir un plan de gestión de datos de investigación.	12	27%	7	22%	19	25%
Total	45	100%	32	100%	77	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación, todo apoyo al investigador respecto a la gestión de sus datos de investigación debe incluirse en la etapa de difusión (capacitación e inducción).

Se determinó que los investigadores, si su organización implementará un repositorio de datos de investigación para la gestión de los mismos, esperarían más de una opción de apoyo. Los apoyos expuestos son 5. El total de opciones de apoyo que esperarían los 25 investigadores son 77 (las opciones de apoyo se repiten en promedio 3,08 veces). Los 16 investigadores de la UPC en total utilizan 45 opciones de apoyo, a un promedio de 2,81 opciones esperadas por investigador, mientras que los 9 investigadores de la UNMSM en total utilizan 32 opciones de apoyo, a un promedio de 3,55 opciones de apoyo por investigador. El 25% (19 investigadores) esperarían como opción de apoyo que en el repositorio se incluya un plan de gestión de datos de investigación. El 22% (17 investigadores) esperarían como opción de apoyo que en el repositorio se incluya cursos de capacitación o entrenamiento. El 19% (15 investigadores) esperarían como opción de apoyo que se incluya un gestor de información especializado en el procedimiento de datos de investigación. El 18% (14 investigadores) esperarían como opción de apoyo en la elección de estándares, licencias de acceso abierto y meta datos. Y por último el 16% (12 investigadores) esperarían como opción de apoyo en directrices o políticas sobre la gestión de datos de investigación.

Como se observa las opciones de apoyo solicitadas son variadas, sin embargo, los investigadores solicitan en su mayoría planes de gestión de datos de investigación.

Tabla 79 Distribución de frecuencias: preferencia del responsable para hacerse cargo del archivo de datos de investigación de los investigadores de CS de la UNMSM - UPC

19. ¿Quién le gustaría que sea el responsable del archivo de sus datos de investigación en un repositorio de datos?

Respuesta	UPC		UNMSM		Ambos	
	fi	hi	fi	hi	fi	hi
Centro de TI	3	14%	5	36%	8	22%
Gestor de la Información	3	14%	5	36%	8	22%
Yo mismo	3	14%	1	7%	4	11%
Yo mismo, pero acompañado de un experto en datos de investigación (data curator)	13	59%	3	21%	16	44%
Total	22	100%	14	100%	36	100%

De acuerdo al modelo del ciclo de vida de los datos de investigación, la responsabilidad sobre los archivos de los datos de investigación de los docentes investigadores está asociada a la etapa de preservación.

Se determinó que a los investigadores les gustaría que el responsable de su archivo de datos de investigación en un repositorio sea más de una persona o institución. Los responsables propuestos son 4. El total de responsables de archivo de los 25 investigadores son 36 (los responsables de archivo se repiten en promedio 1,44 veces). Los 16 investigadores de la UPC en total recurrirían a 22 responsables de archivo, a un promedio de 1,37 responsables de archivo por investigador, mientras que los 9 investigadores de la UNMSM en total recurrirían a 14 responsables del archivo, a un promedio de 1,55 responsables de archivo por investigador. Al 44% (16 investigadores) le gustaría que el responsable de su archivo fuera el mismo, pero acompañado de un experto en datos de investigación (data curator). Al 22% (8 investigadores) le gustaría que el responsable de su archivo fuera un Centro de TI. A otro 22% (8 investigadores) le gustaría que el responsable de su archivo fuera un Gestor de la Información. Y al 11% (4 investigadores) le gustaría ser el mismo el responsable de su archivo.

Se observa que los investigadores estarían dispuestos a colocar sus datos de investigación en un repositorio de datos, siempre y cuando estén acompañados de un experto o gestor de la información.

5.4. Conclusiones de la encuesta realizada en la UNMSM y la UPC

Se elaboraron las conclusiones sobre los resultados de las encuestas realizadas en: la UNMSM a 9 investigadores docentes; y en la UPC a 16 investigadores docentes, haciendo un total de 25 investigadores docentes.

Las conclusiones sobre los resultados de la encuesta realizada en la UPC, se dividieron de acuerdo a los bloques de estudio: Bloque I: información general; Bloque II: uso de datos de investigación; Bloque III: organización de datos de investigación; y Bloque IV: almacenamiento de datos de investigación.

En el Bloque I se plantearon las preguntas del 1 al 3 del cuestionario, con el objetivo de ubicar el lugar donde el investigador realiza su investigación, determinar el origen de sus datos, la línea o líneas de su investigación, y de donde provienen los fondos de su financiamiento.

De acuerdo a los resultados obtenidos podemos resaltar que de las 41 investigaciones en marcha la mayoría tienen más de una línea de investigación, y que la mayoría de los investigadores realizan sus investigaciones en la institución donde se desempeña como docente, esta condición les permite utilizar los recursos administrativos, tecnológicos, científicos y financieros que la Universidad gestiona, así también refleja que las líneas de investigación en las cuales trabajan son muy diversas, sin embargo las que más sobresalen a las que se dedican los investigadores son Educación en Ciencias de la Salud y Epidemiología y se descubrió que hay investigadores que reciben financiamiento de más de una agencia.

En el Bloque II se plantearon las preguntas del 4 al 9 del cuestionario, con el objetivo de determinar el uso que el investigador da a sus datos, cómo, dónde y cuánto tiempo conserva sus datos de investigación, si los comparte, da acceso y cómo y qué medios utiliza para acceder a los datos de investigación de otros investigadores, con el fin de determinar cómo los investigadores pueden acceder a los datos.

En referencia a los resultados obtenidos en este bloque se obtiene un nivel muy alto en la conservación de datos ya que el 100% de investigadores conserva sus datos de investigación y que también existe una buena disposición de parte de los investigadores para compartir sus datos de investigación con los demás, pero no lo hace debido a que esta medida está condicionada a la presencia de un medio o una plataforma confiable en la cual se pueda intercambiar los datos. Así mismo, se encontró que los investigadores están dispuestos a conceder el acceso a sus datos de investigación si se los solicitan. Sin embargo se halló que los investigadores utilizan diversos dispositivos electrónicos y medios informáticos, para compartir sus datos de investigación y que tienen ciertas dificultades, para reutilizar los datos de investigación de otros investigadores, por falta de información sobre los beneficios de la reutilización. Así mismo se encontró que los investigadores están dispuestos a aplicar a licencias que les permitan proteger sus datos de investigación. Se descubrió que los investigadores acceden a catálogos y bases de datos especializados.

En el Bloque III se plantearon las preguntas del 10 al 13 del cuestionario, con el objetivo de determinar los aspectos organizativos sobre el registro, formatos, técnicas de recolección y el tipo de dato de investigación que utilizan los investigadores con el propósito de que estos datos puedan ser reutilizados.

En lo que se refiere a los resultados obtenidos en este bloque, se identifica que los investigadores utilizan diversos medios y formatos para registrar y describir sus datos de investigación, así como también más de una técnica de recolección de datos, siendo la encuesta la técnica de recolección que más se utiliza; un grupo menor utiliza la técnica experimental, pocos la entrevista, y muy pocos las historias clínicas. Se halló que los investigadores utilizan más de un tipo de datos de investigación. La mayoría utiliza datos numéricos, encuestas y resultados de la medida de instrumentos y que no utilizan estándares apropiados que puedan servir para la reutilización de los datos.

En el Bloque IV se plantearon las preguntas del 14 al 19 del cuestionario, con el objetivo de determinar las condiciones, volumen, tipo de almacenamiento, disponibilidad y como almacenar los datos de investigación, con el fin que estos puedan ser preservados.

En relación a los resultados obtenidos en la encuesta, determinamos que los investigadores requieren más de una condición para enviar sus datos de investigación a un repositorio de datos; asimismo, que la mayoría de investigadores utiliza menos de un Tera como espacio de almacenamiento para guardar sus datos de investigación y que los investigadores utilizan más de un tipo de almacenamiento para depositar sus datos de investigación; Se encontró que los investigadores tienen más de una razón para que sus datos de investigación estén disponibles y preservados; y que la condición que más resalta es que mejora la transparencia de los resultados; después de esta razón de acuerdo a la importancia detallada son: la disponibilidad, duplicación de esfuerzos, validación de resultados y uso de resultados; y finalmente se encontró que los investigadores, si bien están de acuerdo en que su organización implemente un repositorio de datos de investigación, solicitarían varias opciones de apoyo, estas son según su importancia: la inclusión de un plan de gestión de datos de investigación, cursos de capacitación o entrenamiento, la inclusión de un gestor de información especializado en el procedimiento de datos de investigación, ayuda para la elección de estándares, licencias de acceso abierto y metadatos, y por último, ayuda a identificar lineamientos o políticas sobre la gestión de datos de investigación. En relación a la persona responsable del archivo de sus datos de investigación en un repositorio institucional, se descubrió que a la mayoría, les gustaría ser ellos mismos, pero acompañado de un experto en datos de investigación (data curator); a otro grupo le gustaría que fuera un Centro de TI; algunos que fuera un gestor de la Información, y a muy pocos les gustaría ser ellos mismos los responsables de su archivo de datos de investigación.

En resumen, al analizar las respuestas obtenidas por los investigadores de acuerdo a los objetivos específicos del estudio (uso, organización y almacenamiento) y los ítems que engloban cada uno de los bloques en relación a la gestión de datos de investigación basados en el ciclo de vida de los datos, se constata que existe en los docentes investigadores la necesidad de información de calidad respecto a la gestión de los datos de investigación y a cada una de las etapas del ciclo de vida de los datos. Para satisfacer esta necesidad el modelo propuesto incluye la etapa de difusión que incluye las distintas formas de apoyo a los investigadores, descrito en el capítulo 3 del presente estudio.

5.5. Beneficios que aporta la propuesta

Los beneficios de la propuesta van más allá de los beneficios que reporta la inclusión de un elemento importante dentro del ciclo de vida de los datos como es la difusión, un elemento clave para la captación y fidelización de los docentes investigadores. En ese sentido, la propuesta en sí misma es un beneficio, de acuerdo a las siguientes consideraciones:

1. Es el paso inicial para el intercambio de conocimientos entre investigadores, intercambio sujeto a políticas claras que incrementen la cantidad de investigaciones; dinamicen la generación de nuevo conocimiento; y promuevan la interacción entre investigadores.
2. Permite el ahorro en el esfuerzo de generar datos de investigación cuando estos ya existen y pueden ser utilizados, sin costo alguno, lo que hace que el investigador sea más eficiente en función al tiempo y a los resultados.
3. Contribuye a generar confianza en el docente investigador para la reutilización de datos de investigación generados por otros investigadores, evitando barreras que impiden mejorar la calidad de las investigaciones.
4. Garantiza los derechos de autor y la propiedad intelectual y evita cargas económicas y administrativas que afectan la eficacia de la actividad investigadora.
5. Utiliza una plataforma informática adecuada que permite la conservación y almacenamiento de los datos de investigación, por líneas de investigación, utilizando formatos y metadatos apropiados para facilitar la generación de los datos y su posterior utilización y reutilización.

CAPÍTULO VI: CONCLUSIONES GENERALES

6.1. Conclusiones respecto al Objetivo General

Se elaboró el Modelo Propuesto para la Gestión de Datos de Investigación basada en el Ciclo de Vida de los Datos y a la vez se validó el modelo a través de la aplicación de una encuesta. Sobre los resultados obtenidos se concluye que los investigadores generan sus datos de investigación y los conservan muchas veces sin compartir con otros investigadores. La gestión de los datos de investigación requiere que la creación o generación de datos por los investigadores se dé dentro de un proceso de localización y recolección de los datos generados por los investigadores, en la UNMSM y UPC, y que la recepción de estos datos se realice en formatos adecuados para el registro de los metadatos. También se concluye la necesidad de contar en la propuesta con la acción de evaluar, debido a que los datos, así recibidos, requieren de ser interpretados analizados y seleccionados de acuerdo a políticas previamente establecidas en el plan de gestión de datos. Otra conclusión asociada al objetivo general tiene origen en el resultado que indica que los docentes investigadores buscan dar seguridad a sus datos de investigación, para ese fin se incluye en la propuesta la acción de preservar, mediante esta acción de utilidad, el dato es seleccionado, pero también puede ser desechado, trasladado a otro archivo, depósito o centro de datos de acuerdo a una política establecida sobre su vigencia y utilidad. También se concluye que los docentes investigadores no sólo buscan un lugar seguro para conservar sus datos de investigación sino también acciones de apoyo y orientación para mantener almacenados de manera temporal sus datos de investigación en sus propios equipos, en la propuesta se incluye la acción de almacenar, para realizar todas estas actividades e incluir otras como: la corrección de errores, limpieza de datos, validación de formatos, protección de datos, copias de seguridad, condiciones de envío de los datos de investigación y la automatización de los procesos para garantizar la calidad del dato en el tiempo. Otra conclusión importante sobre los resultados del estudio, es que los docentes investigadores no tienen información sobre cómo acceder a los datos de investigación de otros investigadores, así tenemos que la acción de acceder, es crítica, incluye la capacitación de los docentes investigadores para que puedan realizar indexaciones, reutilizar datos de otros investigadores, establecer licencias y obtener

permisos de acuerdo a las políticas establecidas. En el estudio también se estableció una conclusión sobre que los docentes investigadores no reutilizan los datos de investigación generados por otros investigadores, por ello la necesidad de incluir en la propuesta la acción de reutilizar los datos de investigación, pero para ello los datos de investigación deben estar debidamente almacenados y disponibles, el objetivo de la reutilización es promover la generación de nuevas investigaciones y por lo tanto a generar nuevo conocimiento, esta acción se puede realizar con más eficiencia debido a las licencias establecidas para los datos abiertos. Otra conclusión importante del estudio es que los docentes investigadores tienen potencial para realizar investigaciones en varias líneas de investigación, por tal razón, se incluyen en la propuesta la acción de transformar que se enfoca en la reutilización de datos originales para transformarlos o crear nuevos datos que generen nuevos resultados y que a su vez generan un nuevo ciclo de vida de los datos. Por último, se concluye que actualmente los docentes investigadores no tienen la información suficiente sobre como: compartir sus datos de investigación con otros investigadores; reutilizar otros datos de investigación; acceder; preservar; y almacenar adecuadamente sus datos de investigación, por estas razones se incluye en la propuesta la acción de difundir, que está constituido por un programa de difusión que se enfoque en la difusión de los beneficios de compartir y reutilizar los datos de investigación, así como en la captación y reclutamiento de docentes investigadores del Área de Ciencias de la Salud, para que el programa de difusión tenga éxito se requiere de un plan de actividades de inducción.

6.2. Conclusiones respecto a los Objetivos Específicos

1. Respecto a la determinación sobre el uso de los datos de investigación generados por los investigadores. Se determinó que los docentes investigadores usan sus datos de investigación sin compartirlos. Esto es, que no están disponibles y por lo tanto acceder a ellos es difícil. Para evitar esta barrera, se debe considerar la necesidad de establecer, protocolos específicos, claros y precisos, especialmente para las primeras etapas del ciclo de vida como son: la generación y conservación de los datos de investigación. Estos protocolos deben ser elaborados de manera que faciliten compartir los datos de investigación con otros investigadores. También se debe tomar en cuenta que el uso de los datos de investigación depende

de dónde están localizados y como han sido recolectados y si los formatos para el registro de estos datos es el adecuado. En ese sentido, debe también considerarse como necesaria la etapa de evaluación, con el propósito de que los datos sean interpretados y analizados. Así se concluye, que existe la necesidad de contar en la propuesta con la etapa de evaluar, que utilice procedimientos que posibiliten que los datos sean interpretados, analizados y seleccionados, con la finalidad de dar la seguridad, que buscan los investigadores, a sus datos de investigación. También se concluye que en la elaboración de la propuesta se incluya la etapa de preservar, con procedimientos que den seguridad al dato y la posibilidad de conservar los datos de investigación o desechar aquellos que no sean útiles.

2. Respecto a cómo están organizados los datos de investigación. Se determinó que los datos de investigación que generan los docentes investigadores, no cuentan con un sistema de descripción de metadatos estandarizados que asegure el acceso, la recuperación y reutilización de los datos por otros investigadores. Así concluimos que, en el marco de la organización de los datos de investigación, se debe definir el proceso de localización, recolección y recepción de los datos en formatos adecuados que faciliten el registro de los metadatos que viabilice el acceso a los datos y consecuentemente a la información. De esta manera, se presenta la necesidad de establecer dentro del ciclo de vida, la etapa de acceder, con el propósito de simplificar y controlar el acceso a los datos, así como, establecer licencias y permisos idóneos. En este punto cabe hacer notar que los docentes investigadores no tienen información sobre cómo acceder a los datos de investigación de otros investigadores, lo que corrobora la necesidad de la etapa de difundir. También podemos concluir que la organización de los datos de investigación permitirá a los docentes investigadores desarrollar varias líneas de investigación. En este sentido, es también necesario incluir la etapa de reutilización, porque siendo varias las líneas de investigación en el área de las Ciencias de la Salud, en estas se generan datos que pueden reutilizarse fácilmente en una u otra línea de investigación. De este modo, se determina la necesidad de incluir también en la propuesta la etapa de la transformación. Esta etapa, directamente asociada a la creación de nuevos datos, a partir de datos originales,

permite generar nuevos resultados, lo que conduce a la generación de un nuevo ciclo de vida de los datos.

3. Respecto a la determinación sobre el proceso de almacenamiento. Se determinó que el proceso de almacenamiento de los datos de investigación, es realizado por los docentes investigadores de las universidades en diferentes dispositivos, formatos y medios. Los datos de investigación también son conservados en sus propios equipos de cómputo, lo que reduce su integración a un contexto mayor y la posibilidad de incluirse en un plan de preservación digital. También se concluye que los docentes investigadores no sólo buscan un lugar seguro para conservar sus datos de investigación sino también acciones de apoyo y orientación para mantener almacenados de manera temporal sus datos de investigación en sus propios equipos, contando con el apoyo de un especialista o experto en el manejo de datos de investigación. Por tales razones, es necesario incluir en la propuesta la etapa de almacenar y difundir. Almacenar para realizar actividades como: corrección de errores, limpieza de datos, validación de formatos, protección de datos, copias de seguridad, envío de los datos y automatización de los procesos. Y difundir, con el propósito de que los docentes investigadores reciban información suficiente sobre el uso y los beneficios de compartir, reutilizar y preservar sus datos de investigación, agregándoles valor académico y científico a los mismos.

REFERENCIAS BIBLIOGRÁFICAS

1. Albrecht, Michael (2010). *Esto of a data repository for a long-running physics experiment*. Tesis para optar el grado de Master of Science in Computer Science and Engineering, Graduate Program in Computer Science and Engineering, Notre Dame, Indiana, USA.
2. Arquero, R., Marcos, G., (2014). El portal de datos abiertos de la Unión Europea: análisis y evaluación. *Revista General de Información y Documentación* 24-1, 99-118. Recuperado de:
http://dx.doi.org/10.529/rev_RGID.2014.v24.n1.45384
3. Asociación Peruana de Facultades de Medicina (2016). *Miembros*. Recuperado de: <http://www.aspefam.org.pe/miembros.htm>. Lima, ASPEFAM.
4. Attarian, Ioanna María (2012). *An Analysis and reasoning framework for Project Data Software Repositories*. Tesis para optar el título para optar el grado de Master of Applied Science in Electrical and Computer Engineerin, Waterloo, Ontario, Canada.
5. Balanchandra, Sujana (2003) “Data Repository for Autonomously guided behibele using ASP.NET”. Tesis para optar el grado de Master of Science in the otros Department of Mechanical, Industrial University, India.
6. Barrios, E., Torrén, R., Nuñez, L. y la Cobarocación de Lago (2011). Implementación de un Repositorio de Datos Científicos usando Dspace. *E-Colabora*, 1(2):102-117. Recuperado de:
<http://publicaciones.renata.edu.co/index.php/RCEC/article/view/50/pdf>.
7. Ball, A., UKOLN, & Bath, U. of. (2009). *Scientific Data Application Profile Scoping Study Report*. Recuperado de:
<http://www.ukoln.ac.uk/projects/sdapss/papers/ball2009sda-v11.pdf%5Cnhttps://drive.google.com/file/d/0BwVuBK4FRW2zYjVKVUdQcGIRdUk/edit?usp=sharing>.
8. Carol, T., Dalton, E. d., Allard, S., Frame, M., Pjesivica, I., Birch, B., ... Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions Among Scientists Worldwide. *PLoS ONE*, 10(8), 1–24. Recuperado de:
<http://doi.org/10.5061/dryad.1ph92>.

9. Commission, E. (2016). *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*. Retrieved from https://www.google.com.pe/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjKovKlyJHQAhWC5iYKHUEOBrqQFggmMAA&url=http%3A%2F%2Fec.europa.eu%2Fresearch%2Fparticipants%2Fdata%2Fref%2Fh2020%2Fgrants_manual%2Fhi%2Foa_pilot%2Fh2020-hi-oa-data-mgt_en.pdf&usg=AFQjCNFDzPikw_7rRA6qZrf6EUqYrxSQUw
10. Chou, Howard (2005) “BioDig: Architecture for Integrating Heterogeneous Biological Data Repositories Using Ontologies”. Tesis para optar el grado of Master of Engineering in Electrical Engineering and Computer Science. Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Techonoly, Massachusetts, USA.
11. Deis, R (2001). La investigación básica en América Latina. *Acta Andina*, 9(1-2), 65-67. Retrieved from 73(2). Recuperado de: http://sisbib.unmsm.edu.pe/bvrevistas/acta_andina/v09_n1-2/investigacion_basica.htm.
12. España. Ministerio de Industria, E. y T. E. (2015). *Cómo publicar datos abierto de manera rápida y sencilla (con Ckan)*. Retrieved from https://www.google.com.pe/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwi5l_OAyZHQAUIyyYKHcBFC1QQFgggBMAA&url=http%3A%2F%2Fdatos.gob.es%2Fsites%2Fdefault%2Ffiles%2Fguia-publicar-opendata-24h_v05.pdf&usg=AFQjCNEt58uKtIPG1QWdAUI0_OfT0GLW4g. Madrid.
13. European Research Council. (n.d.). *Guidelines on the Implementation of Open Access to Scientific Publications and Research Data in Projects supported by the European Research Council under Horizon 2020* (Vol. Horizon 20).
14. Federer, L. M., Lu, Y.-L., Joubert, D. J., Welsh, J., & Brandys, B. (2015). Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff. *PloS One*, 10(6), e0129506. Recuperado de: <http://doi.org/10.1371/journal.pone.0129506>.
15. Ferrer-Sapena, A., Peset, F., and Aleixandre-Benavent, R. (2011). Acceso a Los Datos Públicos y su reutilización: Open Data y Open Government. *El Profesional de la Informacion* 20(3): 262,263. Retrieved from

<http://elprofesionaldelainformacion.metapress.com/openurl.asp?genre=article&iid=doi:10.3145/epi.2011.may.03> (June 21, 2014).

16. Fuente, G. B. de la, & Jesús Robledano Arillo. (2012). *Una aproximación al data curation y el rol del bibliotecario en su implantación*. Universidad Carlos III de Madrid. Retrieved from:
https://www.google.com.pe/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjE1-WDypHQAhXCKyYKHcoUDB4QFggbMAA&url=http%3A%2F%2Farchivo.uc3m.es%2Fbitstream%2Fhandle%2F10016%2F16251%2FTFM_%2520aproximacion_perez_2012.pdf%3Fsequence%3D1&usg=AFQjCNF_R2NWyKmxzxvC99NegECAcns40g
17. Gómez, Nancy-D., Méndez, E., Hernández-Pérez, T. (2016) Datos y metadatos de investigación en ciencias sociales y humanidades: una aproximación desde los repositorios temáticos de datos. *El profesional de la Información* 25(4).
18. Hernández-Pérez, T., García-Moreno, M. (2013). Datos Abiertos Y Repositorios de Datos: Nuevo Reto Para Los Bibliotecarios. *El Profesional de la Informacion* 22(3): 259,260.
19. Humboldt-Universität zu Berlin. (2014). *A supplement to the Humboldt-Universität zu Berlin Research Data Management Policy*. Berlin. Retrieved from:
https://www.google.com.pe/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwiO44fSy5HQAhUIZCYKHVeNAGcQFggbMAA&url=https%3A%2F%2Fwww.cms.hu-berlin.de%2Fde%2Fdl%2Fdataman%2Fhu-rdm-guidelines%2Fat_download%2Ffile&usg=AFQjCNGiggRLzJIdMI8_iUCctYVQLoIFQQ
20. J Thorp. (2015). *Questionnaire National Research Data Survey*.
21. La Referencia visibilizando la ciencia. Red Federada de Repositorios Institucionales de Publicaciones Científicas. Recuperado de:
<http://lareferencia.redclara.net/rfr/>
<http://lareferencia.redclara.net/rfr/sites/default/files/LAReferenciaTresPaginas.pdf>

22. Ley N° Ley N° 30035. Ley que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto (05 de junio de 2013) Lima: Congreso de la República. Recuperado de:
<http://portal.concytec.gob.pe/images/stories/images2013/portal/areas-institucion/dsic/ley-30035.pdf>.
23. Luzi, D., Ruggieri, R., Biagioni, S., & Schiano, E. (2013). Data sharing in environmental sciences: A survey of CNR researchers. In *International Journal of Grey Literature* (Vol. 9, pp. 69–81). Retrieved from:
http://www.opengrey.eu/data/70/01/82/GL14_Luzi_et_al_2013_Conference_Preprint.pdf.
24. Martínez, J., (2001). *Nacimiento del SPARC Europe* [en línea]. Revista Española de Documentación Científica 24, 4. Retrieved from:
<http://redc.revistas.csic.es/index.php/redc/article/viewFile/73/134>
25. McNeill, Katherine. (2007) "Interoperability Between Institutional and Data Repositories: a Pilot Project at MIT." International Association for Social Science Information Service and Technology. Retrieved from:
<http://hdl.handle.net/1721.1/66981>.
26. Melero, R. (2016). Jornada Open in action: los datos de investigación, un paso más hacia la Ciencia abierta [diapositiva]. España, Universidad Politécnica de Valencia. Retrieved from <http://www.action.openaccessweek.org/>.
27. Millán-González, L., Saorín, T., Ferrer-Sapena, A., Aleixander-Benavent, R. and Peset, F. (2013). Gestión de Datos de Investigación: Infraestructuras para su difusión. *El Profesional de la Información* 22(5): 417-419. Retrieved from:
<http://elprofesionaldelainformacion.metapress.com/openurl.asp?genre=article&iid=doi:10.3145/epi.2013.sep.06>.
28. Ministerio Federal alemán para la Educación y la Investigación, (2009), *Mapa de Investigación-Perú*. Bonn, Oficina Internacional del BMBF, Alemania
29. Muñante, D. (2010). *DSpace aplicado al repositorio institucional RedLIEDS (Red Latinoamericana sobre Industrias Extractivas y Desarrollo Sostenible)*. Tesis para optar el título profesional de Ingeniero de Sistemas, Facultad de Ingeniería de Sistemas e Informática. Universidad Nacional Mayor de San Marcos (UNMSM). Retrieved from:
<http://cybertesis.unmsm.edu.pe/handle/cybertesis/3169>.

30. Nakano Silvia; Azlirevich Paola (24 de octubre 2016) Datos primarios de investigación científica: diagnóstico, estrategias y desafíos para la promoción de la ciencia abierta en América Latina [video conferencia](..min...s.). Recuperado de [http:// www.youtube](http://www.youtube)
31. OECD (2015) Gobierno Abierto en América Latina, OECD Publishing, París.DOI:10.1787/9789264225787-es. Recuperado de:
<http://www.oecd.org/gov/gobierno-abierto-en-america-latina-9789264225787-es.htm>
32. OECD (2015) Gobierno Abierto en América Latina, OECD Publishing, París.DOI:10.1787/9789264225787-es. Recuperado de:
<http://www.oecd.org/gov/gobierno-abierto-en-america-latina-9789264225787-es.htm>
33. OECD (2015) Evaluación y recomendaciones, en Gobierno Abierto en América Latina, OECD Publishing, París.DOI:10.1787/9789264225787-3-es. Recuperado de:
<http://www.oecd.org/gov/gobierno-abierto-en-america-latina-9789264225787-es.htm>
34. OECD (2015)Aprovechar el potencial para los datos abiertos en América Latina, in en Gobierno Abierto en América Latina, OECD Publishing, París.DOI:10.1787/9789264225787-7-es. Recuperado de:
<http://www.oecd.org/gov/gobierno-abierto-en-america-latina-9789264225787-es.htm>
35. OPEN AIRE Guidelines for Data Archives. Retrieved from <https://guidelines.openaire.eu/en/latest/data/introduction.html>
36. Pacheco-Romero, J., Huerta, D., Galarza,C., et.al (2012). Producción científica de los Institutos y Centro de Investigación de la Facultad de Medicina de la Universidad Nacional Mayor de San Marcos. *Anales de la Facultad de Medicina*. 73(2). 147-51. Retrieved from:
http://www.scielo.org.pe/scielo.php?pid=S1025-55832012000200011&script=sci_arttext.
37. Perú. Consejo Nacional de Ciencia, Tecnología e Innovación (2016). *Directrices para el procesamiento de información en los repositorios institucionales*. Retrieved from [alicia..concytec.gob.pe](http://alicia.concytec.gob.pe). Lima, CONCYTEC

38. Perú. Consejo Nacional de Ciencia, Tecnología e Innovación (20 de noviembre de 2016). Portal de Datos Abiertos. Recuperado de: <http://www.datosabiertos.gob.pe>. Lima, CONCYTEC.
39. Perú: Presidencia del Consejo de Ministros (2016). Seminario internacional: Desarrollando innovación en Transparencia: datos abiertos en el marco de los estándares de la OCDE.. Recuperado de <http://sgp.pcm.gob.pe/wp-content/uploads/2016/09/Datos-Abiertos-en-el-marco-de-los-estandaresOCDE.pdf>. Lima, Presidencia del Consejo de Ministros.
40. Reglamento de la Ley 30035. Ley que regula el Repositorio Nacional Digital de Ciencia, Tecnología e innovación de Acceso abierto (25 de enero de 2015) Lima: Presidencia del Consejo de Ministros. Recuperado de: <http://spij.minjus.gob.pe/normas/textos/240115T.pdf>.
41. Risco de Dominguez, G. (2014) Diseño e implementación de un currículo por competencias para la formación de médicos. Rev. Peru.Med.Exp. Salud Pública. 31:572-81.
42. Rogova, Ermir (2010). *Treatment of imprecision in data repositories with the aid of KNOLAP*. Tesis para optar el grado de Doctor en Filosofía, Universidad de Westminster, Londres. GB.
43. Rubio, F (2009) “La estructura organizacional en Centros de Investigación, Desarrollo e Innovación; una aproximación a la experiencia internacional. Tesis para optar el grado de Doctor en Administración, Facultad de Contaduría y Administración, Santiago de Querétaro, México.
44. Salaverry, Oswaldo. El inicio de la educación médica moderna en el Perú: La creación de la Facultad de Medicina de San Fernando. Acta méd. peruana [online]. 2006, vol.23, n.2 [citado 2016-11-21], pp. 122-131 . Disponible en: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1728-59172006000200013&lng=es&nrm=iso. ISSN 1728-5917.

45. Sancho R.(2005). *Directrices de la OCDE para la obtención de indicadores de Ciencia y Tecnología*. Madrid, España. Retrieved from:
http://www.micit.go.cr/encuesta/docs/docs_tecnicos/ocde_directrices_para_indicadores_ciencia_y_tecnologia.pdf.
46. Sapena, Antonia Ferrer and Fernanda Peset Mancebo. 2013. “Gestión de Datos de La Investigación.” Universitat Politècnica de València. Retrieved for:[https://riunet.upv.es/bitstream/handle/10251/36053/Tesina](https://riunet.upv.es/bitstream/handle/10251/36053/Tesina_final_de_estudios_Máster_Oficial_CALSI.pdf?sequence=1) final de estudios Máster Oficial CALSI.pdf?sequence=1).
47. Starr, J., Ammann, N., Ashton, J., Barton, A., Elliott, J., & Jacquemot-Perbal, M. (2015). *DataCite Metadata Schema for the Publication and Citation of Research Data*. Retrieved from:
<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:DataCite+Metadata+Scheme+for+the+Publication+and+Citation+of+Research+Data#1>
48. Torres-Salinas, D., Robinson-García, N., Cabezas-Clavijo, Á. (2012). Compartir los datos de investigación en ciencia: introducción al data sharing. *El Profesional de La Informacion*, 21(2),174,175. doi:10.3145/epi.2012.mar.08.
49. Torres-salinas, D., Martín-Martín, A., & Fuente-Gutiérrez, E. (2014). NOTAS Y EXPERIENCIAS / NOTES AND EXPERIENCES Analysis of the coverage of the Data Citation Index – Thomson Reuters : disciplines , document types and repositories, 37(1), 1–6.Torres-Salinas, D., Robinson-García, N., Cabezas-Clavijo, Á. (2012). Compartir los datos de investigación en ciencia: *Informacion*, 21(2), 173–184. doi:10.3145/epi.2012.mar.08.

ANEXOS

Anexo 1 Glosario

ADMS: Asset Description Metadata Schema. Esquema para describir archivos semánticos.

AGA: Alianza para el Gobierno Abierto.

ALCs: Países de América Latina

ALICIA: Repositorio Nacional Digital de Ciencias, Tecnología e Innovación de Acceso Abierto.

CAFME: Comisión de Acreditación de Facultades o Escuelas de Medicina Humana

COAR (Confederación de Repositorios de Acceso Abierto): Asociación internacional con más de 100 miembros y socios de todo el mundo que representan bibliotecas, universidades, instituciones de investigación, financiadores del gobierno y otros. COAR reúne a la comunidad de repositorios y a las principales redes de repositorio para crear capacidad, alinear políticas y prácticas y actuar como una voz global para la comunidad de repositorios

CONCyTEC: Consejo Nacional de Ciencia, Tecnología e Innovación Tecnológica.

CI+DI+I: Centros de Investigación ,Desarrollo e Innovación

CKAN (Comprehensive Knowledge Archive Network): Sistema o plataforma de aplicación web que sirve para la publicación de datos de código abierto.

CONI: Consejo Nacional de Investigaciones del Perú.

CSV: Representar datos en forma de tabla; la columna se separan por comas y las filas por salto de página.

DAG: Datos de Gobierno Abierto

DATA CITE: esquema de metadatos para la identificación precisa de un recurso con el fin que sea recuperado.

DATA CURATOR: Curador de datos que realiza las actividades que implican la selección, registro, clasificación y almacenamiento de los datos con el fin que estos sean reutilizados.

DATA CURATION: Conjunto de acciones para el mantenimiento y utilización de los datos digitales y los resultados de investigación.

DC (Dublin Core): es un formato de metadato definido sobre la base de un consenso internacional. El conjunto de elementos de Metadatos Dublin Core define quince elementos para una descripción y descubrimiento simple del recurso.

DKAN: Software de código abierto que sirve para la publicación de datos abiertos.

DMP: Plan de Gestión de datos.

DRIVER (Digital Repository Infrastructure Vision for European Research): Proyecto financiado por la Unión Europea para crear una infraestructura de servicios y de datos de investigación a partir de la agregación de los contenidos de los repositorios abiertos europeos y adaptado a nivel internacional.

EBC: Educación basada en competencias.

DCAT: Data Catalog Vocabulary. Vocabulario diseñado para facilitar la interoperabilidad entre los diferentes catálogos de datos publicados en la Web.

DOIs: Identificadores de objetos digitales

EDAG: Estrategias de gobierno Abierto

EUROSTAT: Oficina de Estadística de las Comunidades Europeas.

EUROVOC: tesoro multilingüe y multidisciplinario que abarca la terminología de los ámbitos de actividades de la Unión Europea, con especial hincapié en las labores parlamentarias. Esta disponible en 23 lenguas oficiales de la Unión Europea.

FINCyT: Fondo para la Innovación, la Ciencia y la Tecnología.

FONDECYT: Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica.

GOBIERNO ABIERTO: Nueva forma de hacer gestión pública.

LATINDEX: Sistema de Información sobre las revistas de investigación científica, técnico-profesionales y de divulgación científica y cultural que se editan en los países de América Latina, el Caribe, España y Portugal.

LICENCIA CREATIVE COMMONS (CC): Es un permiso para usar o compartir obra en el ámbito digital (sobretudo internet), pero no equivale ni reemplaza a la inscripción en el Registro de Propiedad Intelectual. Más bien, ambos instrumentos se complementan y refuerzan la protección de obra.

LINKED DATA: forma parte de la familia de estándares de la Web Semántica desarrollados por el World Wide Web (W3C). Este término hace referencia a un conjunto de buenas prácticas para la publicación e interconexión de datos estructurados en la Web, introducidas por Tim Bernes-Lee y que han llegado a ser como los principios de linked data.

LSID: Life Science Identifiers, es un modo para nombrar y localizar piezas de información en la web. Esencialmente, un LSID is un identificador único para algún tipo de data y el protocolo LSID especifica un modo estandarizado de localizar data.

NIH: National Institutes of Health

NSF: National Science Foundation

OCDE: Organización para la Cooperación y el Desarrollo Económico)

OAI : Open Archives Initiative, es una iniciativa para desarrollar y promover estándares interoperables que apuntan a facilitar la diseminación eficiente del contenido para los archivos que contienen los contenidos digitales.

OAI Protocol for Metadata Harvesting (OAI-PMH): es un protocolo ligero de la cosecha para el intercambio de metadatos entre servicios.

OGP: Open Government Partnership

OLAP: Acrónimo en inglés de “procesamiento analítico en línea”. Es una solución utilizada en el campo de la llamada “Inteligencia Empresarial (Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos.

OPENAIRE Guidelines: Directrices de Open Aire para publicar metadatos en repositorios de datos.

OPENDOAR: Directorio de repositorios en acceso abierto.

OWL: es el acrónimo de Web Ontology Language, un lenguaje de marcado que se utiliza para publicar y compartir datos utilizando ontologías en la Web.

RDF (Research Definition Framework): Modelo de representación para la descripción de recursos siguiendo los fundamentos de la Web Semántica.

RDF DATA CUBE: recomendación de la W3C que permite relacionar datos multidimensionales, como las estadísticas en la Web, de tal manera que permite vincular conjuntos de datos relacionados y conceptos.

RDF VIRTUOSO: base de datos de motor híbrido que combina diferentes funcionalidades (servidor de aplicaciones Web, base de datos virtual, servidor de archivos, etc.)

RED3DATA: Registro global de repositorios de datos de investigación que cubre los diferentes disciplinas académicas.

SPARQL (SPARQL Protocol and RDF Query Language): es un lenguaje estandarizado para la consulta de datos RDF de la World Wide Web Consortium, (W3C).

XML: Lenguaje general de etiquetado que define un conjunto de reglas para la codificación de documentos a través de modelos específicos legibles y procesables.

XHTML: Extensible hypertext markup language

W3C: World Wide Web Consortium

Anexo 2 Relación de Facultades de Medicina: Asociación Peruana de Facultades de Medicina (ASPEFAM)

Universidad Nacional Mayor de San Marcos; Universidad Peruana Cayetano Heredia; Universidad Nacional de Trujillo; Universidad Nacional San Luis de Gonzaga de Ica; Universidad Nacional de San Agustín (Arequipa); Universidad Nacional Federico Villareal; Universidad Nacional de Piura; Universidad Nacional Pedro Ruiz Gallo (Lambayeque); Universidad San Martín de Porres; Universidad Católica de Santa María (Arequipa); Universidad Nacional del Centro del Perú (Huancayo); Universidad Ricardo Palma; Universidad Privada Antenor Orrego (Trujillo); Universidad Privada San Pedro (Chimbote); Universidad Cesar Vallejo (Trujillo); Universidad Peruana los Andes (Huancayo); Universidad Nacional de la Amazonía Peruana (Iquitos); Universidad Nacional del Altiplano (Puno); Universidad Privada de Tacna; Universidad Nacional de Cajamarca; Universidad Científica del Sur; Universidad Privada San Juan Bautista; Universidad Nacional José Faustino Sánchez Carrión (Huacho); Universidad Peruana de Ciencias Aplicadas; Universidad Nacional de San Antonio Abad del Cusco; Universidad Nacional Daniel Alcides Carrión; Universidad de Piura; Universidad Nacional de Ucayali; Universidad Particular de Chiclayo; Universidad Alas Peruanas; Universidad Católica Santo Toribio de Mogrovejo.

Anexo 3 Documentos de trabajo, Diseño del Modelo propuesto

Propuesta de diseño de un modelo para la gestión de datos de investigación			
Variables de la investigación	Ítems	Propuesta: gestión de datos de investigación en base al ciclo de vida de los datos	
Bloque I: información general Considerar el diseño de un modelo para la gestión de datos de investigación en cada Universidad.	1. Lugar de investigación	Creación de los datos	Universidades 1. UNMSM 2. UPC
	2. Líneas de investigación	Creación de los datos	Ciencias de la salud 1. Salud Pública y Gestión en Salud
	3. Origen de fondos	1. Creación de los datos	2. UNMSM 3. UPC 4. Entidades privadas 5. Gobierno nacional
Bloque II: uso de los datos de investigación: Considerar el diseño de un modelo de gestión de datos de investigación que permita preservar, compartir, facilitar el intercambio, dinamizando el acceso a datos de investigación de otros investigadores, en formatos compatibles con los dispositivos electrónicos y medios informáticos que usan los investigadores, que promueva la reutilización con políticas claras respecto a las licencias para proteger los derechos de autor y propiedad intelectual.	4. Uso de datos	Análisis de los datos: evaluación y selección de los datos para su preservación	1. Disponibilidad 1.1. Generación 1.2. Localización 1.3. Acceso abierto 2. Indexación temática 3. Técnicas de recolección: Manual; Automático 3.1 De acuerdo al proceso de obtención 3.1.1 Experimentales 3.1.2 Observacionales 3.1.3 Resultados de la medida de instrumentos 4. Tipos de datos 1.1. Según el formato 1.1.1. Datos numéricos 1.1.2. Datos de encuestas

	5. Disponibilidad de datos	Preservación de los datos	1. Preservación 1.2.Conservación 1.3.Preservación 1.4.Mantenimiento 1.5.Desecho o traslado 2.Tiempo 1.6.Política sobre la duración del ciclo de vida 3.Seguridad 3.1 Backup 3.2 Servicio de Nube
	6. Acceso a los datos	Acceso y uso de los datos	1. Acceso a los datos 1.1. Distribuir 1.2. Compartir 1.3. Controlar 2.
	7. Medios de acceso	Acceso y uso de los datos	1. Medios de acceso 1.1.Infraestructura o plataforma/captura de datos
	8. Uso de datos de otros investigadores	Acceso y uso de los datos	1. Acceso a datos de otros investigadores 1.1. Dispositivos electrónicos 1.2.Repositorio de datos 1.3.Web personal o institucional 1.4.Material complementario
	9. Reutilización, licencia	Reutilización de los datos	1. Reutilización 1.1.Creación de nuevos resultados 2. Licencias 2.1. Permisos de reutilización

Bloque III: organización de datos de investigación: Considerar como parte del diseño del modelo, los estándares internacionales para registrar los datos de investigación adecuadamente, elegir los formatos de almacenamiento idóneo, las técnicas de recolección de datos y el tipo de datos de investigación a preservar.	10.Registro	Reutilización de los datos	1. Registro 1.1. Esquema de metadatos/ DataCite 1.2. Esquema de metadatos/ OpenAireGuidelines for Data Archive
	11.Formatos	Reutilización de los datos	1. Formatos 1.1. Hojas de cálculo XLS, ODS, CSV, SAS 1.2. Documentos de Texto DOC, TXT
Bloque IV: almacenamiento de datos de investigación: Con los hallazgos obtenidos, se justifica el desarrollo de una plataforma informática que cumpla con la mayoría de condiciones requeridas por los investigadores en lo que se refiere al espacio de almacenamiento, tipo de almacenamiento, disponibilidad, preservación, transparencia, validación y uso de resultados, incluyendo un plan de gestión de datos, un gestor de información especializado o experto (data curator), con el acompañamiento necesario para la elección de estándares, licencias de acceso abierto, capacitación en el uso de	12. Condiciones/almacenamiento	Almacenamiento de los datos	1. Condiciones/almacenamiento 1.1. Soporte gestionable para su transferencia y almacenamiento 1.2. Subdominio propio 1.3. Interoperabilidad 1.4. Calidad de los datos 1.4.1 Lineamientos y políticas para la gestión de datos (confiabilidad y seguridad) 1.5. Supervisión 1.5.1 Roles y responsabilidades del administrador e investigador
	13. Espacio de almacenamiento	Almacenamiento de los datos	1. Espacio de almacenamiento 1.1. capacidad de la plataforma
	14. Tipo de almacenamiento	Almacenamiento de los datos	1. Tipo de almacenamiento 1.1. Servicio en la nube 1.2. Repositorio institucional
	15. Razones de disponibilidad	Almacenamiento de los datos	1. Razones de disponibilidad

formatos con directrices y políticas claras y precisas.			
	16.Apoyo	Almacenamiento de los datos	1. Apoyo 1.1 Gestor de información especializado en el procedimiento de datos de investigación 1.2 Plan de gestión de datos
	17.Responsable	Almacenamiento de los datos	1. Investigador acompañado de un gestor de la información